THE ROLE OF p107 IN MYOGENIC STEM CELL DNA METHYLATION

KYRA FINE

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN

BIOLOGY

YORK UNIVERSITY

TORONTO, ONTARIO

APRIL, 2024

Abstract

Myogenic stem cells have the potential to either self-renew or differentiate into muscle fibers. These stem cell fate choices are influenced by metabolic changes which affect the epigenetic landscape, including DNA methylation which alters gene expression patterns. Because we have found that a mitochondrial function of p107, a transcriptional co-repressor, is involved in regulating SC metabolism and fate decisions, we investigated whether p107 influences the mitochondrial DNA methylation signature. For this we used a portable long read sequencer MinION (Nanopore), to sequence and compare the mitochondria DNA methylation pattern of wildtype and p107 genetically deleted (p107KO) C2C12 myogenic cell line. We found that p107KO mitochondrial DNA had decreased methylation levels for OXPHOS genes compared to wild type. Thus, suggesting that p107 might suppress ND4 gene expression. We also compared a novel method of mtDNA isolation to three existing methods in order to determine the most efficient approach for sequencing.

Acknowledgements

I would like to express my deepest appreciation for Professor Anthony Scimè, who has guided and mentored me throughout the research process, providing invaluable advice and support. I would also like to express my deepest gratitude to Professors Sebastian Magierowski and Ebrahim Ghafar-Zadeh, who provided the sequencing materials and resources that made this research possible, as well as their expertise and insight on computing and informatics. I also received critical assistance from my lab mates Vicky Shah and Justin Hsiung, who were kind enough to let me use the p107KO C2C12 cell cultures they had generated. I also owe thanks to Professor Peter Backx, who sat on my masters' supervisory committee and provided valuable advice during the planning of my research project, and to Professor Logan Donaldson, who joined the supervisory committee during the thesis writing process. I would also like to thank my family, for supporting me throughout the process and for acting as a sounding board.

Table of Contents:

Abstract……..ii

Acknowledgements……..……iii

Table of Contents…….……..…...iv

List of Tables…….…..……..……..v

List of Figures……..……vi

List of Abbreviations……..……..…….……vii

Chapter 1: Introduction

1.1. Mitochondria control of DNA methylation…….……..1

1.2. Mitochondria DNA…….…….……...…..3

1.3. Mitochondrial genome methylation…….……..……...6

1.4. Impact of DNA methylation…….…….....……7

1.5. Methods for assessing DNA methylation……..……...9

1.6. MinION DNA methylation sequencing……..10

1.7. Rbl1 (p107) function in mitochondrial metabolism……12

1.8. Evidence for p107 involvement in mtDNA methylation……15

1.9. Rational…….……..15

1.10. Hypothesis…….16

1.11. Objectives……..17

Chapter 2: Materials and Methods……18

Chapter 3: Results

3.1. mtDNA sequencing reads……...26

3.2. Mitochondrial alignment……29

3.3. Non-mouse reads……32

3.4. Non-mouse read identification……33

3.5. Read depth over the mitogenome……..36

3.6. Flow Cell Degradation rate……39

3.7. Mitogenome methylation of different DNA isolation methods……..43

Chapter 4: Discussion……..47

Chapter 5: References…….58

Appendix……...69

List of Tables

Table 1. Summary of sequencing results……...28

Table 2. Summary of alignment results…….31

Table 3. Summary of alignment results to contaminating organisms……...34

Table 4. NCBI Blast alignment of Gamma retroviruses against the C3H mouse genome……35

Table 5. differentially methylated sites in a p107WT and a p107KO sample……..44

Appendix 1. differentially methylated sites in *MA* and *MD*……69

Appendix 2. differentially methylated sites in *MC* and *MD*……73

List of Figures

Chapter 1:

Introduction

Figure 1 Cytosine methylation……2

Figure 2. Electron Transport Chain (ETC)……4

Figure 3. The human mitochondrial genome……5

Figure 4. Nanopore mechanism for reading DNA……12

Figure 5. Schematic for p107 mitochondrial function based on Sirt1 activity……14

Chapter 2:

Materials and methods

Figure 6. mtDNA extraction approaches……19

Chapter 3:

Results

Figure 7. Visual representation of mtDNA extraction approaches……27

Figure 8. Read depth over mouse mitochondria genome……37

Figure 9. Sequencing speed of flow cell over the course of the sequencing operation……40

Figure 10. Degradation of the flow cells over the course of the sequencing operation.……42

List of Abbreviations

5hmC	5-hydroxymethylcytosine
5mC	5-Methylcytosine
ADP	Adenosine Diphosphate
ALS	Amyotrophic Lateral Sclerosis
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
Blood and Tissue Kit	DNEasy Blood and Tissue Kit (Qiagen)
bp	Base Pairs
C2C12	A C3H-Derived Mouse Myoblast Cell Line
C33A	Immortalized Human Cervical Cancer Cell Line
C3H	C3H/HeJ *Mus Musculus* Lineage
C5	Carbon 5
Cas9	CRISPR Associated Protein 9
CaSki	Immortalized Human Cervical Cancer Cell Line
CpG	5” Cytosine Followed By 3” Guanine
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
D-Loop	Displacement Loop
DMEM	Dulbecco’s Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DNMT	DNA Methyltransferase
DNMT1	DNA-Methyltransferase 1
DNMT13	DNA-Methyltransferase 1 isoform 3

DNMT13A	DNA-Methyltransferase 1 Isoform 3A
DNMT3A	DNA-Methyltransferase 3A
DNMT3B	DNA-Methyltransferase 3B
dsDNA	Double Stranded DNA
DSS	Dispersion Shrinkage For Sequencing Data
E2F	E2 Transcription Factor
E2F1	E2F Transcription Factor 1
E2F4	E2F Transcription Factor 4
EDTA	Ethylenediaminetetraacetic Acid
ETC	Electron Transport Chain
FAD+	Flavin Adenine Dinucleotide
$FADH_2$	Flavin Adenine Dinucleotide + Hydrogen (H) + Hydrogen (H)
FAST	FAST Analysis of Sequences Toolbox
FAST5	Oxford Nanopore Hierarchical Data Format 5
FASTA	FAST-All
FASTQ	A Variant Of The FASTA Format Which Also Stores Sequence Quality Values
FBS	Fetal Bovine Serum
G1 Phase	Cell Cycle Growth 1 Phase
HCT116	Immortalized Human Colon Cancer Cell Line
HDAC1	Histone Deacetylase 1
HEPES	4-(2-Hydroxyethyl)-1-Piperazineethanesulfonic Acid
HSP1	Heavy Strand Promoter 1

HSP2	Heavy Strand Promoter 2
KO	Knockout
LSP	Light Strand Promoter
MA	Method A
MB	Method B
MC	Method C
MCF10A	Michigan Cancer Foundation Epithelial Cell Line 10A
MD	Method D
ME	Method E
Miniprep Kit	the Qiaprep Spin Miprep Kit (Qiagen)
Mitoepigenome	Mitochondrial Epigenome
Mitogenome	Mitochondrial Genome
mtATP6	Mitochondrially Encoded ATP Synthase Membrane Subunit 6
mtCO1 / COX1	Mitochondrially Encoded Cytochrome C Oxidase 1
mtCO2 / COX2	Mitochondrially Encoded Cytochrome C Oxidase 1
mtDNA	Mitochondrial DNA
mtND1	Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase Chain 1
mtND2	Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase Chain 2
mtND3	Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase Chain 3
mtND4	Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase Chain 4
mtND5	Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase Chain 5
mtND6	Mitochondrially Encoded NADH-Ubiquinone Oxidoreductase Chain 6
NAD^+	Nicotinamide Adenine Dinucleotide

NADH	Nicotinamide Adenine Dinucleotide (NAD) + Hydrogen (H)
NCBI	National Center for Biotechnology Information
NCR	Noncoding Region
ND1	NADH-Ubiquinone Oxidoreductase Chain 1
ND2	NADH-Ubiquinone Oxidoreductase Chain 2
ND3	NADH-Ubiquinone Oxidoreductase Chain 3
ND4	NADH-Ubiquinone Oxidoreductase Chain 4
ND5	NADH-Ubiquinone Oxidoreductase Chain 5
ND6	NADH-Ubiquinone Oxidoreductase Chain 6
nDNA	Nuclear DNA
NGS	Next Generation Sequencing
OS	Operating System
OXPHOS	Oxidative Phosphorylation
PCR	Polymerase Chain Reaction
p-value	Probability Value
Q-score	Phred Quality Score Q
Rb	Retinoblastoma Susceptibility Gene
Rbl1	p107 / Retinoblastoma-Like Protein 1
RG108	N-Phthalyl-L-Tryptophan
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
S Phase	Cell Cycle Synthesis Phase
SAM	S-Adenosylmethionine

SAMC	S-Adenosylmethionine Carrier
sgRNA	Single Guide RNA
Sirt 1	Sirtuin (Silent Mating Type Information Regulation 2 Homolog) 1
ssDNA	Single Stranded DNA
TCA	Tricarboxylic Acid Cycle
TET	Ten-Eleven Translocation Methyl Cytosine Dioxygenase
TET2	Ten-Eleven Translocation Methyl Cytosine Dioxygenase 2
tRNA	Transfer RNA
WIMP	What's In My Pot Epi2ME-Labs Workflow
WT	Wildtype

Chapter 1: Introduction

Mitochondria control of DNA methylation

DNA methylation is an epigenetic process where methyl groups are transferred to DNA, often at the C5 position of the cytosine (Moore et. al., 2013). DNA methylation of bases within gene promoters is known to cause a reduction in transcription by impairing the binding of transcription activators and by recruiting proteins that themselves block transcription factors (Razin and Cedar, 1991, Moore et. al., 2013). Methyl groups are typically transferred to DNA from S-adenosylmethionine (SAM) to the fifth carbon of a cytosine residue to form 5-methylcytosine (5mC) by enzymes called DNA methyltransferases (DNMTs) (Saini et. al., 2017, Moore et. al., 2013) (**Fig. 1**).

An important element that has been shown to influence the methylation of nuclear DNA (nDNA) is the mitochondria (Lopes, 2020, Bellizzi et. al., 2012, Smiraglia et. al., 2008). In adult human blood samples nDNA methylation levels are correlated with mitochondrial genome haplotype (Bellizzi et. al., 2012). Another study on human tumour cell lines showed that nDNA methylation patterns changed in response to reductions in mitochondrial DNA (mtDNA) content and were restored when mtDNA was reintroduced to these cells (Smiraglia et. al., 2008).

The ability of the mitochondria to regulate nDNA methylation levels is possible because mitochondrial activity controls ATP generation (Mookerjee et. al., 2017). Because ATP is a necessary substrate for methylation to occur (Bellizzi et. al., 2012) anything that alters ATP supply, such as changes in mtDNA content and mitogenome haplotype, would impact the availability of ATP for methylation and thus the rate at which methylation can be carried out.

S-adenosyl-methionine (SAM) is the primary methyl group donor in mammalian DNA methylation reactions (Ouyang et. al., 2020). Because the conversion of the SAM precursor methionine to SAM requires ATP, SAM production is limited by mitochondrial metabolism (Ouyang et. al., 2020). The mitochondria also contribute directly to the generation of SAM through the one carbon cycle (Xu et. al., 2021), a process occurring within the mitochondria where the amino acid serine is produced and used to generate folates, which contribute carbon groups necessary for the conversion of methionine to SAM (Li and Hoppe, 2023) (Xu et. al., 2021) (Zhao et. al., 2021). Dietary supplementation with the SAM precursor methionine has been found to increase nDNA methylation rates (Tremolizzo et. al., 2002), supporting the possibility that mitochondrial control of SAM levels may be part of the mechanism through which the mitochondria controls methylation levels. Thus, any mitochondrial mechanisms which regulate ATP production or control the mitochondrial folate cycle, will effect nDNA methylation rates (Morin et. al., 2022).

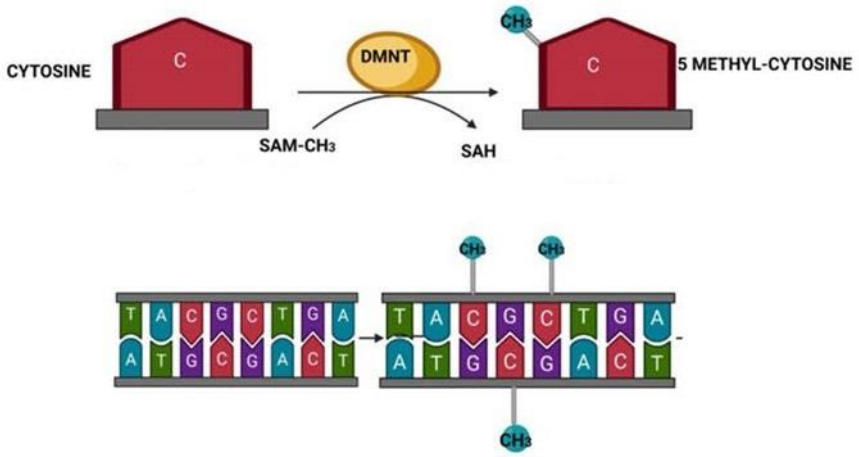


Figure 1. Cytosine methylation. DNMTs catalyze the transfer of a methyl group from S-adenosyl-methionine (SAM-CH_3) to the fifth carbon of a cytosine residue to form 5-methylcytosine (5mC) and produce S-adenosyl-homocysteine **(**Adapted from Milosevic, Arsic, Cvetkovic, and Vucic, 2021).

Mitochondria DNA

Though most of the cell's genetic material is found in the nucleus, the mitochondria contain their own genome (mtDNA), complete with protein coding genes and genetic elements which help regulate mitochondrial replication and gene expression (Annesley and Fisher, 2019) (Nicholls and Minczuk, 2014). mtDNA is circularized, lacks histones, and is maternally inherited (Taanman, 1999). The human mitochondrial genome is 16,569 base pairs long and encodes 2 rRNAs, 22 tRNAs and 13 polypeptides, which code for subunits of the electron transport chain (ETC) involved in oxidative phosphorylation (OXPHOS) (Basu et. al., 2020) (Sharma et. al., 2005) (**Fig. 1**). OXPHOS is the process whereby cells catabolize amino acids, glucose or fatty acid to produce reducing agents in the tricarboxylic acid cycle (TCA), which are oxidized in the ETC in the presence of oxygen to produce ATP from the phosphorylation of ADP (Tang et. al., 2020) (**Fig. 2**). The mitochondrial genome encodes the ETC subunits; mtND1, mtND3, mtND4, and mtND6 for complex I, mtCO1 and mtCO2 for complex IV, and mtATP6 for complex V (Mastroeni et. al.,

2017). mtDNA is believed to be a remnant feature from when mitochondria were single celled organisms that fused with other single celled organisms during the early evolution of eukaryotes (Roger, Muñoz-Gómez, and Kamikawa, 2017). As part of this fusion process many genes that were originally encoded in the mtDNA are believed to have been transferred into the nucleus over time (Brandvain and Wade, 2009).

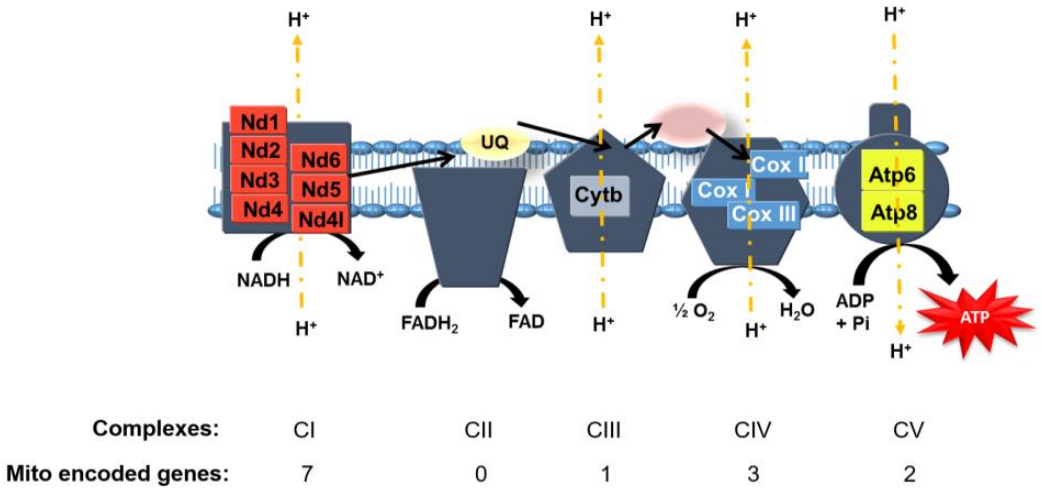


Figure 2. Electron Transport Chain (ETC). This figure depicts the ETC and number of proteins in each of the five OXPHOS complexes which are translated from mitochondrial genes. In the ETC a series of transmembrane electron transporters transfer electrons from NADH and FADH2 through the ETC complexes and eventually to oxygen molecules. The energy of the electron transport pump transfers protons from the mitochondrial matrix to the intermembrane space. The proton gradient allows protons to re-enter the mitochondrial matrix which drives the synthesis of ATP. Each OXPHOS complex is depicted in grey, embedded in the mitochondrial membrane. The mitochondrially encoded proteins contributing to each complex are depicted as coloured boxes containing the name of the protein, located within the corresponding OXPHOS complex. The colour of each protein's box corresponds to the complex it is a component of, with red for complex I, light grey for complex III, blue for complex IV, and yellow for complex V. Adapted from (Bhattacharya, 2017).

The mitochondrial genome is double stranded, with the two strands referred to as the heavy strand and the light strand, based on their buoyancies in caesium chloride density gradients

(D'Souza and Minczuk, 2018). The light strand contains the genes for one ETC subunit (mtND6), and eight tRNAs (Basu et. al., 2020). The Heavy Strand contains the genes for twelve ETC subunits, two rRNAs, and fourteen tRNAs (Basu et. al., 2020). The mitochondrial genome also contains a structure called the D-loop. The D-loop is found in the major noncoding region (NCR) of the mtDNA and is formed by a linear strand of DNA, called 7S DNA (Nicholls and Minczuk, 2014). The D-loop contains the origin of replication for the heavy strand of the mitochondrial genome (Nicholls and Minczuk, 2014) and is involved in regulating the expression of mitochondrially encoded genes (Sanyal et. al., 2018).

During transcription, the entire mitochondrial genome is transcribed as a single unit. The transcription start sites for both the heavy strand and light strand are found in the NCR (Montoya et. al., 1982). The NCR also contains promoter elements upstream of the transcription start sites (Basu et. al., 2020). There is one promoter for the light strand, (LSP) (Bogenhagen, Applegate, and Yoza, 1984), and two promoters for the heavy strand, (HSP1 and HSP2) (Montoya, Gaines, and Attardi, 1983). These promoters drive the expression of the genes encoded in their respective strands (Basu et. al., 2020).

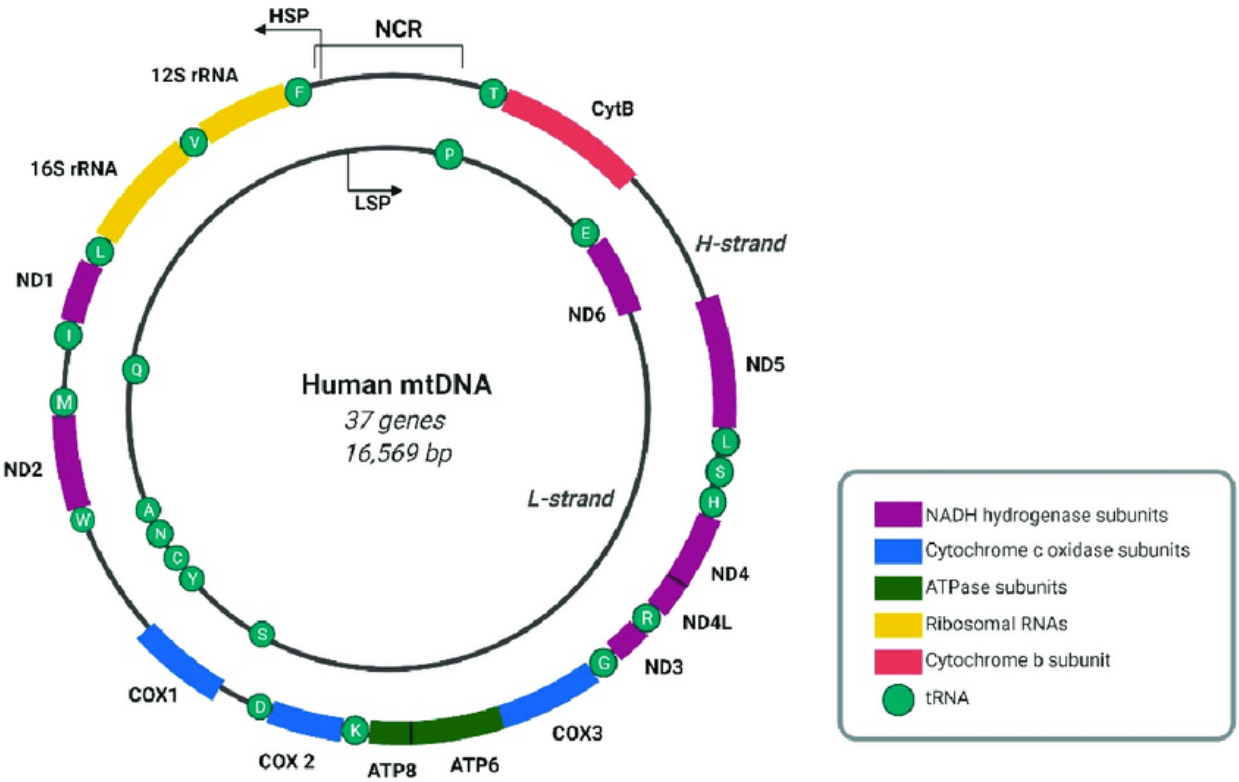


Figure 3. The human mitochondrial genome. Schematic showing the heavy and light strands (H-strand and L-strand), non-coding region (NCR), genes encoding Ribosomal RNAs (orange), gene encoding tRNAs (light green), and genes encoding subunits of the ETC: NADH hydrogenase subunits (purple), Cytochrome c oxidase subunits (blue), ATPase subunits (dark green), and a Cytochrome b subunit (red). The HSP is the promoter region where transcription of the heavy strand begins, and the LSP is the promoter region where transcription of the light strand starts. For the tRNA genes, the letter within the light green circle represents that tRNA's corresponding codon. (Adapted from Jedynak-Slyvka, Jabczynska, and Szczesny, 2021)

Mitochondrial genome methylation

There has been uncertainty in older research about the presence of methylation in the mitochondrial genome and its importance in regulating cellular processes if it is present (Ghosh et. al., 2015). Notably, more current research has found that methyltransferases localize to the mitochondria where they have an important role for the epigenetic regulation of mitochondrial gene expression (Robertson et. al., 2000). A recent study using mouse Oocytes shows that the overexpression of a SAM mitochondria transport protein caused an increase in the amount of SAM entering the mitochondria and potentially methylating mtDNA. The end result was a decrease in

mtDNA gene expression and an increase in reactive oxygen species (ROS), suggesting that OXPHOS was downregulated (Cheng et. al., 2022). These changes suggested that mtDNA methylation inhibited gene expression.

mtDNA methylation might be regulated by a number of methylases. DNA-methyltransferase 1 (DNMT1) isoform 3 (DNMT13) has been found to translocate to the mitochondria (Saini et. al., 2017). Two other methyltransferase proteins, DNMT3A and DNMT3B have also been shown to impact mtDNA methylation status (Chestnut et. al., 2011, Patil et. al., 2019). Different isoforms of DNMT3A are present in the mitochondria in different tissues, with the 78kDa and 100kDa isoform being primarily found in skeletal muscle and neural tissue, respectively (Wong et. al., 2013). DNMT3A has been demonstrated to methylate mtDNA in mouse amyotrophic lateral sclerosis (ALS) models (Chestnut et. al., 2011). Here, DNMT3A expression was found to increase mtDNA methylation in motor neurons that were undergoing apoptosis. When DNMT activity was inhibited with the drug RG108, mtDNA methylation levels returned to normal and apoptosis did not occur, suggesting that DNMT3A activity is critical for the expression of disease phenotype (Chestnut et. al., 2011). Family member DNMT3B has also been demonstrated to be involved in the regulation of mtDNA methylation (Patil et. al., 2019). In the noncancerous human breast epithelial cell line MCF10A, the knockdown of the DNMT3B gene caused a reduction in methylation levels across the mtDNA and an increase in the expression of the ND6 gene (Patil et. al., 2019).

A methyl group modifying DNA enzyme, ten-eleven translocation methyl cytosine dioxygenase 2 (TET2), a non-heme and 2-oxoglutarate-dependent oxygenase, has also been found in the mitochondria (Dzitoyeva, Chen, and Manev, 2013). TET2 oxidizes 5mC methylated bases into 5-hydroxymethylcytosine (5hmC) (Sodiq et. al., 2021). 5hmC may play a role in the regulation

of cellular processes by preventing the inheritance of methylation marks during DNA replication. (Hahn, Szabó, and Pfeifer, 2014). In aged mouse cerebellum neurons, mitochondrial TET levels were found to be lower in young mice, suggesting a link with senescence (Dzitoyeva, Chen, and Manev, 2013).

Impact of mtDNA methylation

Studies have shown that changes to mtDNA methylation can result in alterations to cellular metabolism and to the replication and proliferation of some cell types (Van derWijst et. al., 2017, Feng et. al., 2012, Sanyal et. al., 2018, Sun et. al., 2021, Tubiana, 1989). Similar to nDNA methylation, which alters the expression of nuclear encoded genes (Razin and Cedar, 1991), mtDNA methylation also can alter the expression of mitochondrially encoded genes (Feng et. al., 2012, Sanyal et.al., 2018, Sun et. al., 2021, Sirard, 2019).

In immortalized human cancer cell lines, which have high proliferation rates, such as C33A (cervical cancer) and HCT116 (colon cancer), mitochondrial gene expression has been found to be influenced by mtDNA methylation (van der Wijst et. al., 2017). One study focussing on colorectal cancers found that the D-loop regulatory element within the mitochondrial genome exhibited reduced methylation in the cancer cells relative to healthy cells (Feng et. al., 2012). The importance of controlling mitochondrial gene expression by mtDNA methylation in cancer cells is highlighted by a study using the cervical cancer cell line CaSki (Menga et. al., 2017). The CaSki cells were upregulated for mitochondrial methylation by overexpression of the S-adenosylmethionine carrier (SAMC) gene, which codes for a transport protein that brings SAM into the mitochondria. The resulting increase in methylation suppressed the mitochondrial gene expression of ND6, ND2 and cytochrome b that resulted in the impairment of OXPHOS and a 50% reduction in ATP content in the mitochondria (Menga et. al., 2017). These metabolic changes caused a decrease in the

proliferation of the cancer cells, and the initiation of apoptosis in some of them (Menga et. al., 2017).

The association between mtDNA methylation and increased mitochondrial gene expression has also been found outside of cancer systems (Sanyal et. al., 2018). Blood samples from populations subjected to arsenic exposure, which causes oxidative stress in cells, demonstrated a decrease in the methylation of the D-loop and ND6 gene that were linked to increased mtDNA expression levels and copy number (Sanyal et. al., 2018). Though both the cancer and arsenic studies involve decreased mitochondrial methylation driving increased gene expression, they do not have identical methylation target sites. The D-loop (Feng et. al., 2012, Sanyal et. al., 2018) and ND6 are sites that are methylated in the CaSki cells and the arsenic study, whereas ND2 and cytochrome b sites are targeted only in the CaSki cells (Menga et. al., 2017, Sanyal et. al., 2018).

Mitochondrial methylation is also associated with senescent cells, where increased methylation of the mtDNA encoded COX2 gene was associated with senescence in heart mesenchymal cells (Sun et. al., 2021). The relationship between mtDNA methylation and senescence was found to be causal, because when exogenously produced COX2 protein was introduced into the cells, counteracting the suppression of the gene through methylation, the proliferation rate of the cells increased (Sun et. al., 2021).

Alterations to mitochondrial methylation patterns have been found to have a direct association with disease states (Stoccoro and Coppedè, 2021). Neural tissue samples taken from Alzheimer's patients have demonstrated alterations to mtDNA methylation patterns (Blanch et. al., 2016, Bradley-Whitman and Lovell, 2013). In entorhinal cortex, hippocampus/parahippocampal gyrus and cerebellum neural tissue samples of Alzheimer's patients, mtDNA methylation levels have been shown to be higher than in healthy control samples (Bradley-Whitman and Lovell, 2013)

(Blanch et. al., 2016). There is also evidence that alterations to mtDNA methylation in neural tissue may be associated with ALS. In this case mtDNA methylation induced by the DNMT1 isoform DNMT13A has been shown to be upregulated in human ALS (Chestnut et. al., 2011). In addition, cultured mouse neural cells inhibiting DNMT13A resulted in an impediment to the initiation of apoptosis by procainamide, supporting a role for mtDNA methylation in ALS (Chestnut et. al., 2011).

Methods for assessing DNA methylation

The utilization of methylation sensitive restriction enzymes is a widely used approach to investigate the methylation status of DNA samples. These restriction enzymes cleave DNA at target sequences that are either methylated or unmethylated, depending on the specificity of the enzyme (Zuo et. al., 2009). This results in different cleavage patterns depending on the methylation state of the DNA. After the sample's DNA has been cleaved, it is then amplified by PCR using specific primers to increase the sample density to a sufficient level, so it can be detected without altering the sequence (Hashimoto et. al., 2007). The cleaved fragments can then be assessed through multiple methods. For example, the methylation status of a specific stretch of DNA can be evaluated by gel electrophoresis, as the fragment lengths will differ based on whether they were digested by the methylation sensitive enzyme (Jaffé et. al., 2008).

Another common method for assessing methylation status is bisulfite sequencing. With this method, the DNA sample that is being examined is treated with bisulfite, which converts unmethylated cytosine to uracil (Zuo et. al., 2009, Hayatsu, 2008), and the sample is sequenced, with the presence of cytosine or uracil at a given site indicating whether that site is methylated or unmethylated (Zuo et. al., 2009, Hayatsu, 2008). This sequencing can be performed with any

sequencer that can recognize both cytosine and uracil, and so is amenable for use with next generation sequencing (NGS) (Zuo et. al., 2009, Hayatsu, 2008).

MinION DNA methylation sequencing

Most NGS technologies such as Illumina sequencing assemble a genome by detecting the sequence of short DNA fragments, often around 600 base pairs (bp) long (Amarasinghe et. al., 2020), then connecting them via regions that overlap between fragments from adjacent sections of the genome to construct a full sequence given sufficient read depth (Illumina, Inc., 2010). Long read sequencing, in contrast, reads larger fragments, often over ten thousand bp long (Amarasinghe et. al., 2020). Long read sequencing provides several advantages, such as the sequencing of repetitive regions because longer reads are more likely to extend past these areas and contain not only the repetitive region, but also include distinct sequences that can be identified for assembly (Logsdon, Vollger, and Eichler, 2020). Similarly, when sequencing RNA, long read sequencing can be used to identify different isoforms due to the reads covering multiple exons (Wright et. al., 2022).

The two most prominent long read sequencing technologies are the Oxford Nanopore and PacBio sequencers (Logsdon, Vollger, and Eichler, 2020). In particular, Oxford Nanopore produces two portable sequencing platforms, the MinION and the PromethION 2 Solo (Oxford Nanopore, 2022, Oxford Nanopore, 2023, **Fig. 4**). As a consequence of their portability the MinION and PromethION 2 Solo have lower sequencing throughput than larger sequencers. This is advantageous when studying a smaller genome, such as the mitogenome, which can be sequenced to sufficient read depth for accurate base determination without the need to dedicate resources to sequence an entire cellular genome on a larger throughput sequencer.

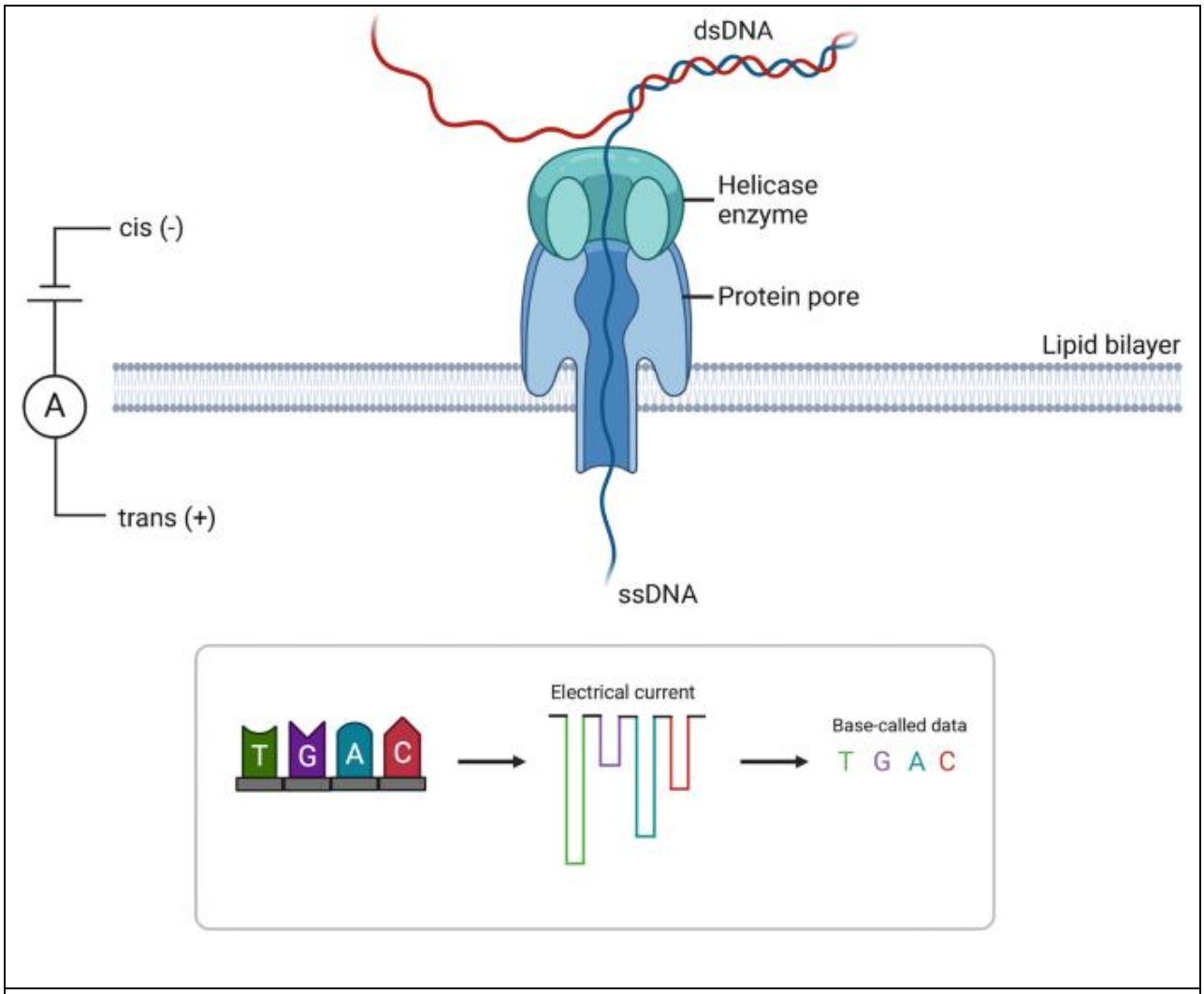


Figure 4. Nanopore mechanism for reading DNA. Double stranded DNA (dsDNA) is denatured by a helicase and single stranded DNA (ssDNA) is passed through a protein pore attached to a lipid bilayer. Nucleotide bases cause differential changes to a constant electrical current by the electrical resistance of each nucleotide to determine the base sequence of the DNA. (Adapted from Beckett and Robson 2023).

Notably, as well as sequencing the genetic code of a sample, Nanopore offers the additional advantage of sequencing the acetylated and methylated DNA bases of that sample (Oxford Nanopore, 2022). Nanopore sequencers can detect these modifications because of the mechanism they use for sequencing (**Fig. 4**). Single DNA strands are run through microscopic pores that are receiving a constant electrical charge (Oxford Nanopore, 2023). The electrical signal across the pore is measured as the DNA strand moves through the pore one base at a time **(Fig. 4)**. Because each nucleotide has a different electrical resistance, the electrical signal across the pore will change

depending on which nucleotide is passing through it (Oxford Nanopore, 2023). Thus, by recording the number and order of each electrical resistance signal that passes through the pore, it is possible to measure the nucleotide sequence of the DNA strand (Oxford Nanopore, 2023). This allows nucleotide identity to be measured for base modifications, as a nucleotide with an attached methyl group will have a different electrical resistance to an unmodified nucleotide of the same type (Flynn et. al., 2022). Any type of base modification can be read if the sequence detection algorithm has been programed with the electrical resistance signals of each nucleotide featuring that base modification (Flynn et. al., 2022). Consequently, base sequences and their modifications can be collected simultaneously from the same sample and as part of the same process (Oxford Nanopore, 2022). Currently, the methylation-detection algorithms for Nanopore sequencing are the best developed and most accurate of the algorithms for detecting different types of DNA modifications (Liu et. al., 2021).

Rbl1 (p107) function in mitochondrial metabolism

Rbl1 (p107) is a member of the retinoblastoma susceptibility gene (Rb) family of transcription factor co-repressors and is most highly expressed during the transition from G1 to S phase in cells undergoing cell division (Wirt and Sage, 2010). It has been shown to cause changes to expression of cell cycle genes through its binding primarily to the E2F4 transcription factor (Araki et. al., 2019, Li et. al., 1997).

Recently, p107 has been shown to suppress mitochondrial gene expression directly through alterations of NAD^+ levels affecting Sirt1 activity (Bhattacharya et. al., 2021) (**Fig. 5**). When the cellular NAD^+/NADH ratio is high in myoblasts, the Sirt1 enzyme prevents p107 from entering the mitochondria. (Bhattacharya et. al., 2021). With p107 unable to localize to the mitochondria it is also unable to interact with mtDNA, resulting in an increase in its gene expression (Bhattacharya et. al., 2021). The suppression of mitochondrial gene expression and the resulting reduction in

OXPHOS activity leads to a reduction in ATP production in the cell from both OXPHOS and glycolysis (Bhattacharya et. al., 2021) (**Fig. 5**). This change to cellular metabolism reduces the cellular proliferation rate and causes the cell to be halted in the G1 phase of the cell cycle (Bhattacharya et. al., 2021). On the other hand, p107 genetically deleted (p107KO) myoblasts exhibit significantly increased OXPHOS levels and proliferation rates relative to control cells (Bhattacharya et. al., 2021) (**Fig. 5**).

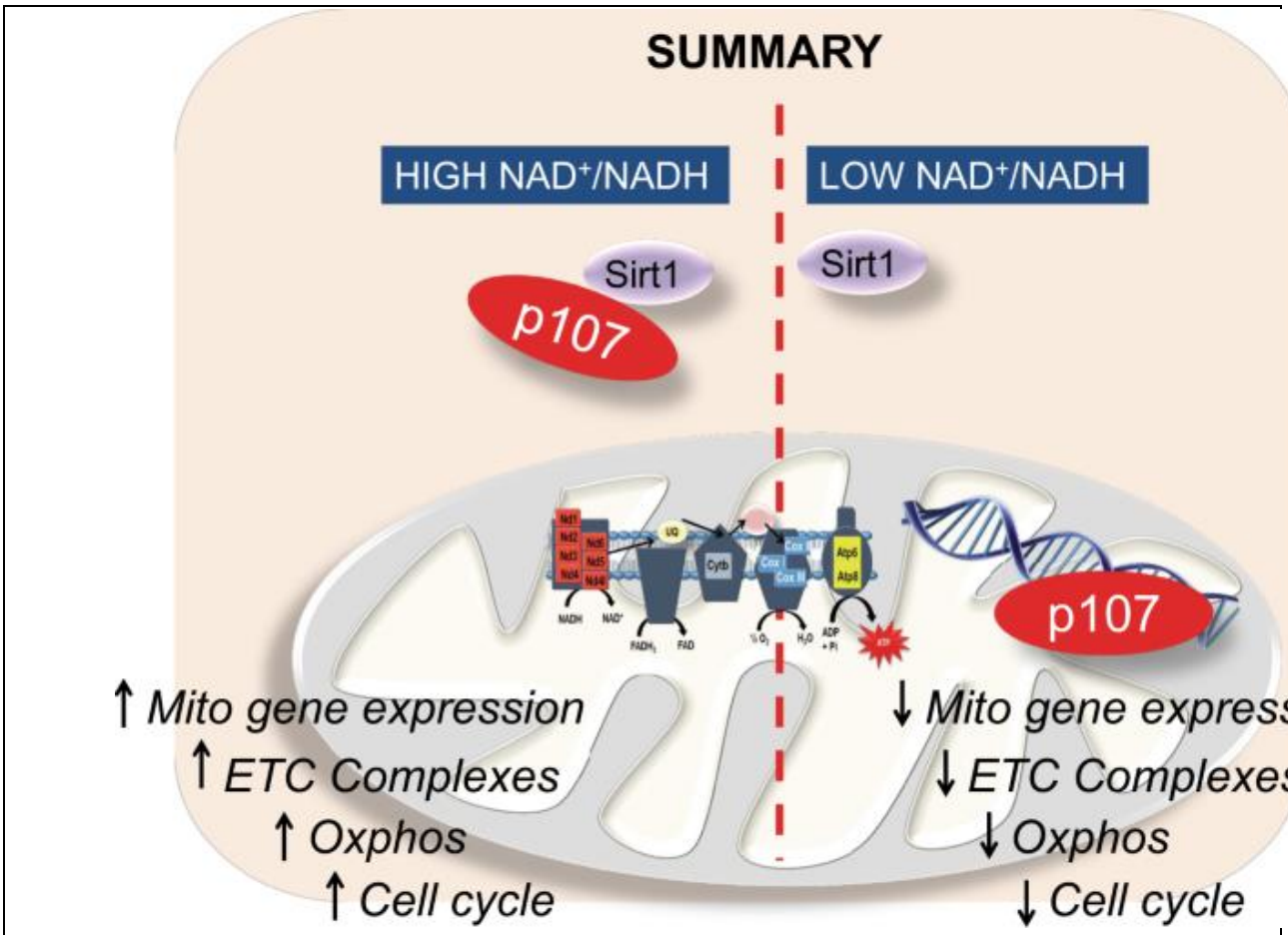


Fig 5. Schematic for p107 mitochondrial function based on Sirt1 activity. When the NAD^+/NADH ratio is low, p107 is unbound from the Sirt1 enzyme, allowing it to access the mitochondrial genome, where it acts to reduce mitochondrial gene expression, reduce the number of ETC complex proteins, and reduce OXPHOS and the cell cycle. A high NAD^+/NADH ratio allows Sirt1 to bind and sequester p107 in the cytoplasm. This causes mitochondrial gene expression to increase, allowing more ETC complexes to be produced, and increasing OXPHOS and cell cycle rate. (Adapted from Bhattacharya et. al., 2021).

Evidence for p107 involvement in mtDNA methylation

Notably, NAD^+ levels within the cell have been shown to impact DNA methylation rates (Ummarino et. al., 2021), potentially implicating p107 in the DNA methylation process. There is also circumstantial evidence suggesting that p107 may be directly involved in methylation processes. As p107 is a member of the Rb family of transcription factors that also include Rb and p130 (Wirt and Sage, 2010), the activity of its family members and their binding partners can lend insight into the potential activity and interacting partners of p107. In this case, Rb has been shown

to form a complex with the E2F1, HDAC1 and DNMT1 proteins, whose methylation targeting has been found to be disrupted in cancer cells (Robertson et. al., 2000). Members of the nuclear Rb/E2F1/HDAC1/DNMT1 complex are known to localize to or have a family member or isoform in the mitochondria (Bhattacharya et. al., 2020, Araki et. al., 2019, Saini et. al., 2017). It is possible that p107 interacts with a DNMT family member because it is a member of the Rb family, is found in the mitochondria and interacts with E2F1 family member E2F4 which is also found in the mitochondria (Araki et. al., 2019, Julian and Blais, 2015, Lee, Bhinge, and Iyer, 2011, Bhattacharya et. al., 2021). Intriguingly, a DNMT1 isoform, DNMT1 isoform 3 (DNMT13), is known to localize in the mitochondria, though its relationship with p107 has not yet been explored (Saini et. al., 2017).

Rational

mtDNA methylation plays a role in several diseases and complications including cancer (Feng et. al., 2012), Alzheimer's (Blanch et. al., 2016, Bradley-Whitman and Lovell, 2013), ALS (Chestnut et. al., 2011), and arsenic poisoning (Sanyal et. al., 2018). mtDNA methylation has also been shown to regulate cellular proliferation (Sun et. al., 2021) and metabolism (Menga et. al., 2017). Thus, because of the importance of mtDNA methylation for human health, the ability to efficiently determine methylation sites would be beneficial for disease screening and research.

The mitochondrial function of p107 presents a potential approach to investigate the mitochondrial epigenetic landscape and test mtDNA extraction methodologies. OXPHOS, which is regulated by p107 (Bhattacharya et. al., 2021), determines the rate of mitochondrial ATP production (Wilson 2017). The conversion of methionine to SAM requires ATP, such that it might impact the availability of SAM for DNA methylation (Ouyang et. al., 2020). Additionally, p107 might interact with mitochondrial methyltransferases (Wirt and Sage, 2010, Robertson et. al.,

2000). Together, this suggests that p107 might influence mtDNA methylation status. Thus, comparison of mitochondria methylation status between p107KO and controls provides a model system to determine mtDNA methylation status and the effectiveness of mtDNA concentration techniques. mtDNA methylation status can be determined efficiently by long read Nanopore sequencing, which can cover the entire mitochondrial genome. (Liu et. al., 2021). Compared to other DNA methylation sequencing methods, Nanopore sequencing allows for the sequencing of both methylation and nucleotides simultaneously and without a chemical nucleotide conversion process, reducing potential sources of error (Oxford Nanopore, 2023).

Hypothesis

Long read Nanopore sequencing can be used to sequence the mitochondrial methylome and elucidate the role of p107 in regulating mtDNA methylation.

Objectives

1. Compare mtDNA concentration and extraction techniques for their sequencing read depth potential.
2. Compare the methylomes of control (p107WT) and p107KO myoblasts to identify genes which are differentially methylated in the absence of p107.

Chapter 2: Materials and methods

Cell lines and culture

The C2C12 myoblast cell line was purchased from the American Tissue Type Culture (ATTC) and grown in Dulbecco's Modified Eagle Medium (DMEM) containing 25 mM glucose supplemented with 10% fetal bovine serum (FBS) and 1% penicillin streptomycin. p107KO cell lines were created and provided by Vicky Shah and Justin Hsiung, two fellow MSc students in the Scimè Lab. To generate p107 genetically deleted (p107KO) cell lines Crispr/Cas9 was used to simultaneously transfect C2C12 cells with 3 pLenti-U6-sgRNA-SFFV-Cas9-2A-Puro plasmids (Applied Biological Materials) each containing a different sgRNA to target p107 sequences (110 CGTGAAGTCATCCAGGGCTT, 156 GGGAGAAGTTATACACTGGC and 350 AGTTTCGTGAGCGGATAGAA).

Mitochondrial isolation

p107KO and control wildtype (p107WT) C2C12 cells were washed in PBS, pelleted, and dissolved in 5 times the packed volume with isolation buffer (0.25 M Sucrose, 0.1% BSA, 0.2 mM EDTA, 10 mM HEPES with 1-5 mg/ml of each pepstatin, leupeptin and aprotinin protease inhibitors), and homogenized in a Dounce homogenizer on ice, three times with the small pestle and three times with the large pestle. The homogenate was centrifuged at 1000 *g* at 4 °C for 10 min. The supernatant was then centrifuged at 14000 *g* for 15 min at 4 °C and the pellet containing the mitochondria was then resuspended in 250µl of QIAprep Spin Miniprep Kit (Qiagen)'s buffer P1 (50 mM Tris·Cl, pH 8.0, 10 mM EDTA, and 100 µg/ml RNase A) for DNA extraction.

Mitochondria DNA extraction

p107WT or p107KO cell cultures were selected for DNA isolation when they reached 60-85% confluency in a 60mm petri dish, for a cell count of approximately $1.9x10^6$-$2.7x10^6$ cells in the growth conditions described above. DNA was extracted using four different methods summarized in **Figure 6**. *Method A (MA)*: p107KO cells had their DNA extracted using the DNEasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol. *Method B (MB)*: Mitochondria were isolated from p107KO cells and their DNA extracted using the DNEasy Blood and Tissue Kit (Qiagen) according to the manufacturer's protocol. *Method C (MC)*: p107KO cells had their DNA extracted using the Qiaprep Spin Miniprep Kit (Qiagen) according to the manufacturer's protocol, which preferentially extracts circularized plasmids (Quispe-Tintaya et. al., 2013). *Method D (MD):* Mitochondria were isolated from p107WT and their DNA extracted using the Qiaprep Spin Miprep Kit (Qiagen) according to the manufacturer's protocol.

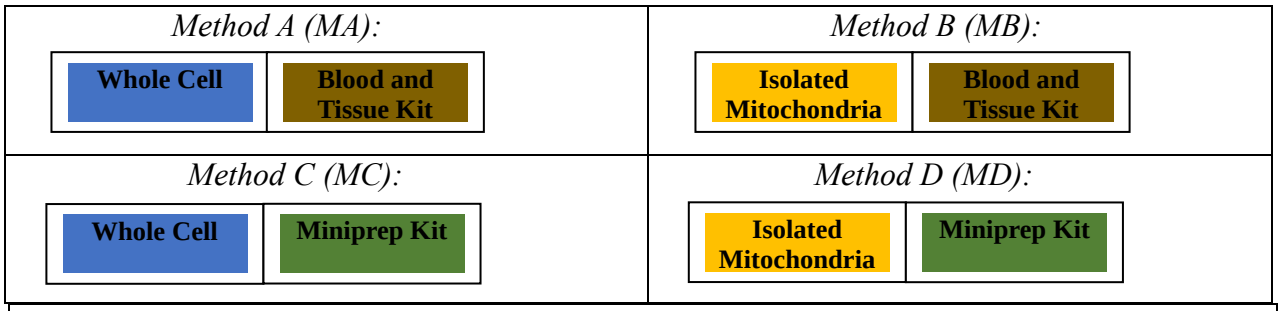


Figure 6. mtDNA extraction approaches. Each mtDNA extraction method is composed of two parts, input material (either whole cell or isolated mitochondria) and DNA extraction method (either Blood and Tissue Kit or Miniprep Kit).

Preparation of DNA libraries for sequencing

For all samples DNA purity and concentration were quantified using a ThermoFisher Scientific NanoDrop 2000 device. After quantification DNA samples were stored at -80°C until needed for sequencing. DNA samples were considered pure enough for use in sequencing if they possessed a 260/280 value greater than 1.8. Library prep was performed according to the manufacturer protocols with the Oxford Nanopore Rapid Sequencing Kit (SQK-RAD004).

DNA sequencing

Oxford Nanopore R9.4 flow cells were loaded in accordance with the manufacturer instructions for the Rapid Sequencing kit. The finished DNA sample library was loaded into each flow cell in a final volume of 75µl. Sequencing was performed using Oxford Nanopore's MinKNOW program using the standard settings for sequencing with an R9.4 flow cell and a SQK-RAD004 kit. Sequencing was performed for 72 hours. Reads were generated during sequencing using Oxford Nanopore's guppy program for base calling, and sequenced reads as well as flow cell signal data were output in FASTQ and FAST5 file formats.

DNA alignment

Alignment of sequenced reads was performed using the guppy base caller with the dna_r9.4.1_450bps_modbases_5mc_hac.cfg configuration. This alignment was performed on three p107KO samples (*MA, MB,* and *MC*), and one p107WT sample (*MD*). During alignment, reads were base called again from the original signal data to determine both nucleotide identity and 5mC methylation status. The base called reads were aligned to the Genome Reference Consortium mouse build 39 mitochondrial reference genome (Bayona-Bafaluy et. al., 2003). Alignment was also performed on the Genome Reference Consortium mouse build 39 whole genome reference genome (Bayona-Bafaluy et. al., 2003). Read coverage across the genome was visualized using Oxford Nanopore's EPI2ME program (Oxford Nanopore, 2023). This was done in order to assess the possibility that nuclear DNA was being misaligned as part of the mitochondria when only the mitochondrial genome was considered.

Comparison of mtDNA isolation methods

The resulting run statistics for each *Method* were then transferred into Libre Office Calc to calculate: the average read length, total number of reads, total number of reads which passed the

quality score cut-off, percentage of reads which passed the quality score cut-off, total number of reads aligning to the mitochondria, number of passed reads aligning to the mitochondria, number of passed reads aligning to no mouse chromosome, percentage of total reads aligning to the mitochondria, percentage of passed reads aligning to the mitochondria, percentage of passed reads aligning to no mouse chromosome, and percentage of passed reads with a known chromosome aligning to the mitochondria.

Adjusting for differences in mtDNA content

As *MD* extracted DNA from WT cells and *MA, MB*, and M*C* extracted DNA from p107KO cells additional analysis was performed to investigate the effects of cell type differences between the samples used on the quantity and concentration of mtDNA extracted. Previous research has shown that p107KO cells possess 5 times the mtDNA copy number levels of p107WT cells generated with a blank CRISPR cassette (Bhattacharya, Shah, and Oresajo, 2021). Thus, the statistical results for *MD* were re-evaluated with the mitochondrial content treated as 5-fold higher, to control for potential differences in mtDNA content. This adjusted version from *MD* is presented in the results as *ME*.

Sequencing Performance

The overall sequencing performance was evaluated using the EPI2ME and EPI2ME-Labs programs (Oxford Nanopore, 2008; Oxford Nanopore, 2023). EPI2ME measured the read depth at each site in the mitogenome, which gives a measure of how accurate we expect the genome and epigenome generated from each method to be (Oxford Nanopore, 2008). It also detects any mitogenomic regions which have been sequenced with either an above or below average level of accuracy. EPI2ME-Labs was used to measure the sequencing speed and number of active flow cell

channels as well as how the sequencing speed and active channel count changed over time during sequencing for each method (Oxford Nanopore, 2023).

Non-mouse read alignment

In order to investigate the origins of passed reads not aligning to the mouse genome, NCBI's BLAST program for alignment (NCBI, 1988) was used to identify a randomly selected read which did not align to the mouse genome. This read most closely aligned to the bacterium *Shigella Flexnari*. Due to its presence in the sample, *Shigella* was used as an example contaminant to investigate whether contaminating DNA was present in each sample, and if so, whether the quantity of contaminating DNA varied between samples. This was tested by aligning each sample to the *Shigella* reference genome *Shigella flexnari* 2a strain 301 (Jin et. al., 2002) using the same method and read quality score cut-offs that were used in the earlier mouse mitogenome alignments. The percentage of passed reads aligning to *Shigella* and the percentage of non-mouse passed reads aligning to *Shigella* were then calculated. Because *Shigella* was present in all samples but not to the same degree, contaminating organisms were determined to be a potential source of error in nanopore based mtDNA sequencing. The extent to which contaminating organisms impacted the sequencing quality of each method was assessed by performing a metagenomic analysis on each sample to determine the identity of each species from which DNA was present in the samples, as well as in what quantity the DNA from each species was present. This analysis was performed using the WIMP (What's In My Pot) workflow from Epi2ME-Labs (Oxford Nanopore, 2023) to generate complete metagenomic alignments of the passed reads from each sample. WIMP can identify a range of viruses and bacteria, but the output of the program does not correctly identify reads aligning to the mouse genome. This occurs because WIMP does not have the mouse genome in its reference database of potential DNA source organisms. This does not disrupt the

metagenomic alignment however, because WIMP instead identifies mouse reads as human due to the similarity between both genomes, and so any reads identified as human can be assumed to instead be from the mouse samples. The validity of this interpretation was confirmed by comparing the number of 'human' aligned reads identified in each sample by WIMP to the number of mouse aligned reads identified by direct alignment of each sample to the mouse genome, which produced similar estimates.

Additional alignments were performed to investigate whether Gamma retrovirus sequences were present in the samples from contamination or as a result of integration into the mouse genome at some point in the evolutionary past. The reference genomes for Mus musculus mobilized polytropic provirus (Evans, L.H., et al) and Murine type C retrovirus (Heinemeyer 2005), the two most common Gamma retroviruses in the sample, were chosen. They were aligned against the reference genome for the C3H mouse strain (Lilue et. al., 2018), which is the mouse strain that the C2C12 cell line was derived from (Yaffe and Saxel, 1977), using the NCBI BLAST program (Altschul et. al., 1990).

DNA methylation reads

Methylation calculation steps, methylation visualization steps, and analysis of methylation steps were performed on both the whole genome and mitochondria-only alignments. The mitochondria-only alignment was chosen for final presentation due to results being consistent across both alignments. mtDNA methylation levels were calculated using Oxford Nanopore's Jupityr Notebook (Kluyver et. al., 2016) based EPI2ME-Labs (Oxford Nanopore, 2023) program. The workflow in the EPI2ME-Labs modified base tutorial was used as the basis for the analysis. The tutorial notebook's code was altered to replace the example datasets with the datasets used for this experiment. These changes were made following the instructions within the tutorial for

replacing the example dataset with user datasets. The notebook code was also altered in one place to correct a coding error that prevented the tutorial code from running properly, as the OS was not loaded into one of the chunks. The calculation of mtDNA methylation levels generated a table of CpG sites across the genome which listed the location of a site, the number of reads covering that site, how many of those reads recorded that site as unmethylated, and how many of those reads recorded that site as methylated.

DNA methylation visualization and analysis

The table of CpG sites and their methylation statuses generated in the previous step was then used to calculate methylation statistics for each sample using Microsoft Excel. The proportion of sites methylated in at least one read, representing methylation distribution, and the overall proportion of individual site reads containing methylation, representing overall methylation level, were both calculated. Methylation data was also visualized using the EPI2ME-Labs modified base tutorial workflow (Oxford Nanopore, 2023). This was done to assess read depth distribution as a bell curve, as well as to visually display the distribution of methylation across the mitochondrial genome and the differences in methylation patterns and levels between the light and heavy strands. Differential methylation between one p107KO sample (*MA, MB,* or *MC*), and one p107WT sample (*MD*) was assessed using the Bioconductor dispersion shrinkage for sequencing data (DSS) package (Park and Wu, 2016) with the R Studio program (RStudio Team, 2020). The mtDNA methylation table generated using EPI2ME-Labs was used as the input for the differential methylation analysis, which was performed according to the DSS user manual to generate a list of differentially methylated regions between the p107WT and each of the p107KO samples. Sample smoothing was used for the DSS analysis because it is required when using only a single replicate for each treatment. Additionally, for the DSS analysis a p-value cut-off of 0.05 was used when

comparing *MA* against *MD* and when comparing *MC* against *MD*. When comparing *MB* against *MD* however, a p-value cutoff of 0.5 was necessary because the single replicate and the older Nanopore library prep kits and flow cells used in this experiment meant that the read depth generated was too low to determine differentially methylated regions with high statistical certainty. Once the differentially methylated regions were determined their positions on the mitochondrial genome were compared to the Genome Reference Consortium mouse build 39 mitochondrial reference genome (Bayona-Bafaluy et. al., 2003) using the NIH genome data viewer to assess the identities of the differentially methylated genes.

Chapter 3: Results

mtDNA sequencing reads

We used four approaches to obtain mtDNA to be sequenced and its methylation status deduced. These were method A (*MA*): p107KO cells had their total DNA extracted using the DNEasy Blood and Tissue Kit (Qiagen), method B (*MB*): mitochondria were first isolated from p107KO myoblast cells and their DNA extracted using the DNEasy Blood and Tissue Kit (Qiagen), method C (*MC*): p107KO cells had their total DNA extracted using the Qiaprep Spin Miniprep Kit (Qiagen), which preferentially extracts circularized DNA including mitochondria (Quispe-Tintaya et. al., 2013) and method D (*MD*): mitochondria were isolated from p107WT myoblast cells and their DNA extracted using the Qiaprep Spin Miprep Kit (Qiagen). Sequencing was performed on the isolated DNA, using Oxford Nanopore's MinION sequencer and MinKnow sequencing program, for a period of 72 hours per sample.

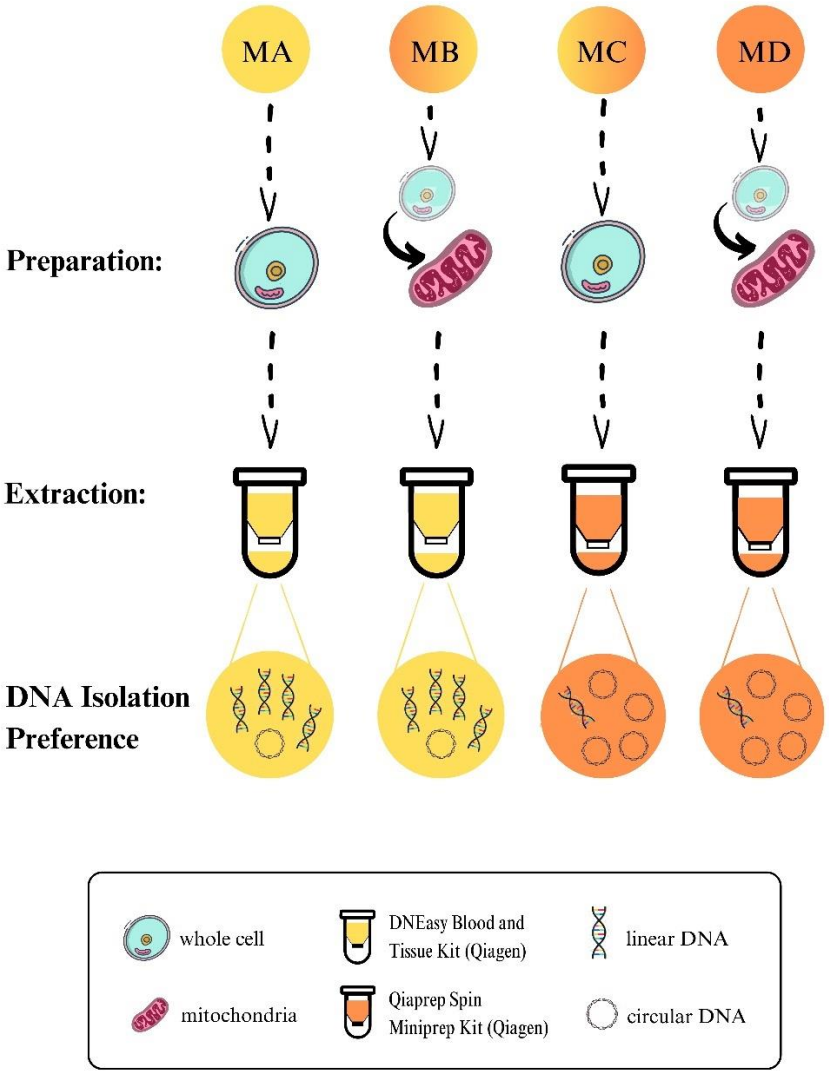


Figure 7. Visual representation of mtDNA extraction approaches. Each mtDNA extraction method is composed of two parts, input material (either whole cell or isolated mitochondria) and DNA extraction method (either the Blood and Tissue Kit, which has no extraction preferences, or the Miniprep Kit, which preferentially extracts circularized DNA).

First, we compared the quality of the DNA for the 4 extraction methods (Fig 7). For Q-scores we used the default Q-score cutoff for Nanopore's Guppy base caller. Thus, we considered only reads with a Phred quality score Q (mathematical representation of the sequencing error probability for that read) of 2 or higher (Delahaye and Nicolas, 2021) (Oxford Nanopore, 2023).

We evaluated the results of each sequencing run for input DNA mass, total and Q-score passing read counts, total and Q-score passing average read length, average length of mitogenome-aligned reads, and percentage of reads passing the Q-score cutoff **(Table 1)**. A fifth analysis, method E (*ME*), was also simulated by recalculating the statistics for *MD* by quintupling the count of mitochondrial-aligned reads in both the total and passed columns. This was done because p107WT cells possess one fifth the mtDNA content of p107KO cells (Bhattacharya, Shah, and Oresajo, 2021).

Table 1. Summary of sequencing results. Statistical summaries of the reads generated during sequencing runs for *MA*, *MB*, *MC*, and *MD*, as well as an adjusted version of the results of run *MD*, presented here as run *ME*. The total quantity of DNA used as input for sequencing, total number of reads, number of reads with a Q-score that is above the experimental Q-score cutoff, average read length, average length of reads with a Q-score above the experimental Q-score cutoff, and percentage of reads which have a Q-score above the experimental cutoff were calculated for each run. The average read length, average passed read length, average passed mitochondrial read length were not separately calculated for sample *ME*.

Sample method	Input DNA mass (ng)	Total reads	Number of passed reads	Average read length (bp)	Average length of passed reads	Percentage of passed reads
MA	400	232253	176278	7619.21	8372.99	75.90%
MB	109.5	94326	41076	5641.60	10073.72	43.55%
MC	166.5	16000	9985	448.14	477.78	62.41%
MD	99	91052	16151	327.19	302.71	17.74%
ME	-	92780	17087	-	-	18.42%

MA possesses a higher number of both total and passed reads than all other samples, at 232253 and 176278 respectively, with *MB* at 94326 and 41076 and *MD* at 91052 and 16151 having similar

counts to each other. *MC* has the fewest at 16000 reads in total and 9985 reads with passing Q-scores (**Table 1)**. *MA* and *MB* also had the largest and second largest average read length, with 7619.21 and 5641.60 bps. But *MB* had the longest average reads with passing Q-scores at 10073.72 bps and the largest difference between the Q-score passing and total average read lengths, with its average Q-score passing read length at 4432.12 bps longer than its average total read length (**Table 1)**. *MC* was the only method to exhibit a reduction in average length between total and passed reads, though with a difference of only 24.48 bps (**Table 1)**.

Together this data might suggest that the additional processing steps taken to concentrate mtDNA in *MB, MC,* and *MD* reduce the total number of reads generated, which is expected when attempting to eliminate nDNA, but also that these steps may fragment DNA, due to the smaller read lengths. The data also suggests that read count and read length are not perfect indicators of each other. In this case, *MB* and *MD* had the highest and lowest average Q-score passing read lengths but did not have the highest and lowest average Q-score passing read counts (**Table 1**). Thus, both should be measured to gain an accurate understanding of the read quality of a sequencing experiment.

Mitochondrial alignment

The reads generated using each method were aligned to the mouse genome to analyze method efficacy at enriching the proportion of mtDNA (**Table 2**). *ME* was analyzed for alignment statistics to assess whether the potential difference in mtDNA content between p107WT and p107KO cells may impact the quality of alignments generated from those cells. Each mtDNA isolation method was analyzed for the total and Q-score passing number of reads which aligned to the mouse mitochondrial genome and mouse nuclear genome, as well as the number of reads which did not align to any part of the mouse genome (**Table 2**). Additionally, the percentage of reads

aligning to the mouse mitochondrial genome was calculated when including all reads in the analysis, as well as the percentage of reads with a passing Q-score which aligned to the mouse mitochondrial genome and the percentage of those passing mouse-aligned reads which specifically aligned to the mitochondria (**Table 2**).

Table 2. Summary of alignment results. Statistical summaries of the alignment of the reads generated during sequencing runs for *MA*, *MB*, *MC*, and *MD*, as well as an adjusted version of the results of *MD*, presented here as run *ME*. For each sample the following were calculated: The total number of reads aligning to the mouse mitochondria, number of reads with a Q-score above the experimental Q-score cutoff which align to the mouse mitochondria, average length of those passing reads which also align to the mitochondria, number of reads with a Q-score above the experimental Q-score cutoff which align to the entire mouse genome, number of reads with a Q-score above the experimental Q-score cutoff which do not align to the mouse genome, percentage of reads with a Q-score above the experimental cutoff which align to the mouse genome, percentage of all reads which align to the mouse mitochondria, percentage of passed reads with a Q-score above the experimental Q-score cutoff which align to the mouse mitochondria, and percentage of mouse-aligned reads with a Q-score above the experimental Q-score cutoff which align to the mouse mitochondria.

Sample method	**Total number of mito reads**	**Number of passed mito reads**	**Average length of passed mito reads**	**Number of passed nuclear reads**	**Number of passed non-mouse reads**	**Percentage of passed reads aligning to the mouse genome**	**Percentage of total reads aligning to the mito**	**Percentage of passed reads aligning to the mito**	**percentage of passed mouse reads aligning to the mito**
MA	443	368	7179.84	174542	1368	99.22%	0.19%	0.21%	0.21%
MB	30	22	5407.05	32056	8998	78.09%	0.03%	0.05%	0.07%
MC	549	510	639.37	5225	4250	57.44%	3.43%	5.11%	8.89%
MD	432	234	1086.47	4893	11024	31.74%	0.47%	1.45%	4.56%
ME	2160	1170	-	4893	11024	35.48%	2.33%	6.85%	19.30%

We found that *ME,* for which the results of *MD* (mitochondrial isolation and plasmid-preference DNA extraction) were adjusted for differences in the mtDNA content of the sequenced

cells, generated the highest number of passed mitochondrial reads at 1170 and *MC,* which used plasmid-preference DNA extraction, generated the second highest number at 510 (**Table 2**). *MA* (no mtDNA concentrating steps) had 368 and *MD (*mitochondrial isolation and plasmid-preference DNA extraction) had 234 passed mitochondrial reads. *MA, MC,* and *MD* all generated mitochondrial read counts in the same order of magnitude and *MB* had the fewest mitochondrial reads with 22 (**Table 2**). *MC* and *MD* generated higher numbers of mitochondrial reads (**Table 2**), while also generating fewer total reads than *MA* (**Table 1**), meaning that more target reads and fewer non-target reads were sequenced. This suggests *MC* and *MD* were more successful than *MA* in isolating more mtDNA for sequencing. Because *ME* generated more mitochondrial reads than either *MB* or *MC*, plasmid-preference DNA extraction and organelle isolation are more effective when used together*,* than they are when only one is used, as in *MB* or *MC.*

As the number of mitochondrial reads in each sample was similar (**Table 2**), the large difference in total read counts between each sample (**Table 1**) might be caused by the number of nuclear reads and reads which do not align to the mouse genome, which differ by a large degree between each sample. The sample which does not seem to hold to this pattern is *MB,* which generated only 22 mitochondrial reads (**Table 2**), despite having a similar number of total reads as *MD*.

Non-mouse Reads

We assessed the non-mouse reads for the different methods of extraction to find the method that provided less contamination. First, we analyzed the percentage of each method's passed reads that aligned to the mouse genome to find the extent of contamination in each method. *MA* had the highest rate with 99.22%, followed by *MB* with 78.09%, with a large reduction for *MC* and *MD* with 57.44% and 31.74%, respectively (**Table 2**). We also determined the number of non-mouse

reads present in each sample. For assessment of contaminating, non-mouse reads, we analyzed each method using only those reads with passing Q-scores. *MD* had the highest number of non-mouse reads at 11024, with *MA MB* and *MC* having non-mouse read counts of 1368, 8998 and 4250 respectively (**Table 2**). This suggests that the mitochondrial isolation and plasmid-preference DNA extraction used in *MD* may have contributed to having the lowest percentage of reads aligning to the mouse genome and the greatest contamination (**Table 2**).

Non-mouse read identification

As there were some reads which did not align with the mouse genome (**Table 2**), we investigated the source organism. For this we performed a metagenomic analysis on each sample using Epi2ME-Labs' WIMP (What's In My Pot) workflow (Oxford Nanopore, 2023) to find the identities and read counts of each contaminating organism in the sample. The reads with passing Q-scores generated from sequencing the mtDNA isolated through each *Method* were used as the input. WIMP outputs a list of all organisms which were the best alignment for at least one read within the sample, as well as how many reads had that organism as their best alignment. For each *Method*, the identities of the five non-mammal organisms with the highest read count and how many reads that organism had as their best alignment are reported in **Table 3**.

Table 3. Summary of alignment results to contaminating organisms. A metagenomic screen was performed on each sample using Oxford Nanopore's Epi2ME-Labs WIMP workflow (Oxford Nanopore, 2023). For *MA*, *MB*, *MC*, and *MD* the five most common contaminating geni are listed, along with how many reads of each genus were present in that sample.

Sample method	**Contaminant 1 identity and read count**	**Contaminant 2 identity and read count**	**Contaminant 3 identity and read count**	**Contaminant 4 identity and read count**	**Contaminant 5 identity and read count**
MA	Escherichia 7349	Salmonella 560	Muvirus 261	Acinetobacter 108	Gammaretrovirus 79
MB	Curtobacterium 1640	Escherichia 1212	Salmonella 246	Gammaretrovirus 33	Mycoplasmopsis 32
MC	Escherichia 2300	Curtobacterium 90	Salmonella 89	Mycoplasmopsis 50	Muvirus 27
MD	Escherichia 4217	Curtobacterium 579	Salmonella 252	Acinetobacter 67	Mycoplasmopsis 40

The list of non-mouse organisms whose DNA was found in the experimental samples included common bacteria such as *Escherichia* and *Salmonella*, as well as Muviruses. Gamma retroviruses were also found, with Mus musculus mobilized endogenous polytropic provirus and Murine type C retrovirus among those identified in *MA* and *MB*. This data indicates that contamination may have occurred while the samples were being prepared and the flow cells were being loaded in the general lab environment.

Table 4. NCBI Blast alignment of Gamma retroviruses against the C3H mouse genome. Mus musculus mobilized endogenous polytropic provirus aligned against C3H in five locations, and Murine type C retroviruses aligned against C3H in one location. For each alignment, the virus being aligned, the C3H contig it aligned to, the query coverage, the percent identity, and the E-value are listed. The query coverage is a measure of the percentage of the viral genome that aligned to the C3H sequence at the location in question. The percent identity is percentage of bases within the aligned segments that are the same in both sequences. The E-value is a measure of the statistical confidence in the alignment being a real similarity and not a result of random chance, with E-values closer to 0 being considered more accurate.

Virus	C3H contig	Query Coverage	Percent Identity	E-Value
polytropic provirus	C3H_HeJ#1#152	2%	94.03%	$1e^{-81}$
polytropic provirus	C3H_HeJ#1#27	2%	93.03%	$3e^{-78}$
polytropic provirus	C3H_HeJ#1#161	2%	93.03%	$3e^{-78}$
polytropic provirus	C3H_HeJ#1#117	1%	90.68%	$7e^{-55}$
polytropic provirus	C3H_HeJ#1#163	1%	91.86%	$1e^{-26}$
Murine type C retrovirus	C3H_HeJ#1#117	4%	76.83%	$1e^{-47}$

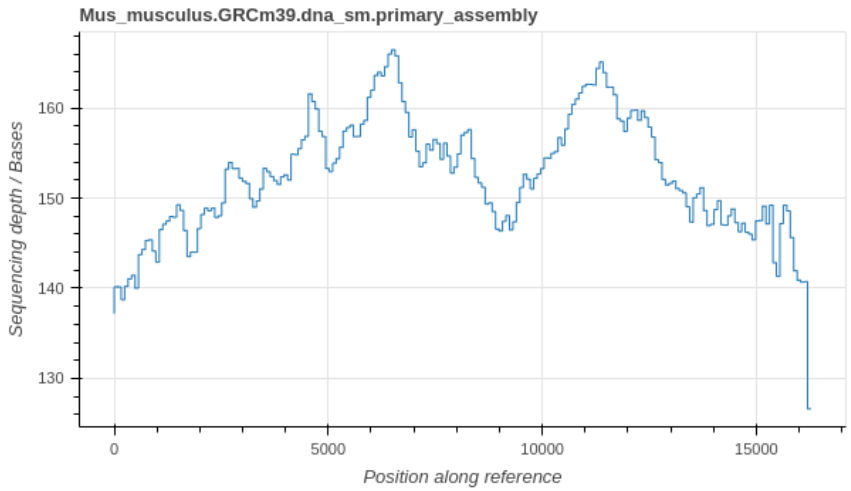
We next assessed if the Gamma retroviruses detected were artefacts within the C2C12 genome rather than contaminants introduced during the sample preparation process. Using NCBI Blast (Altschul et. al., 1990) the reference genomes for Mus musculus mobilized endogenous polytropic provirus (Evans,L.H., et al) and Murine type C retrovirus (Heinemeyer 2005) were both aligned against the C3H mouse reference genome, which is the mouse lineage from which C2C12 cells were derived (Yaffe and Saxel, 1977). Both viruses aligned to C3H with very low E-values of $1e^{-24}$ or lower (**Table 4**). This suggests that portions of the viral genomes may have integrated into the C3H genome, resulting in their misidentification as contaminants by WIMP (**Table 3**).

Read depth over the mitogenome

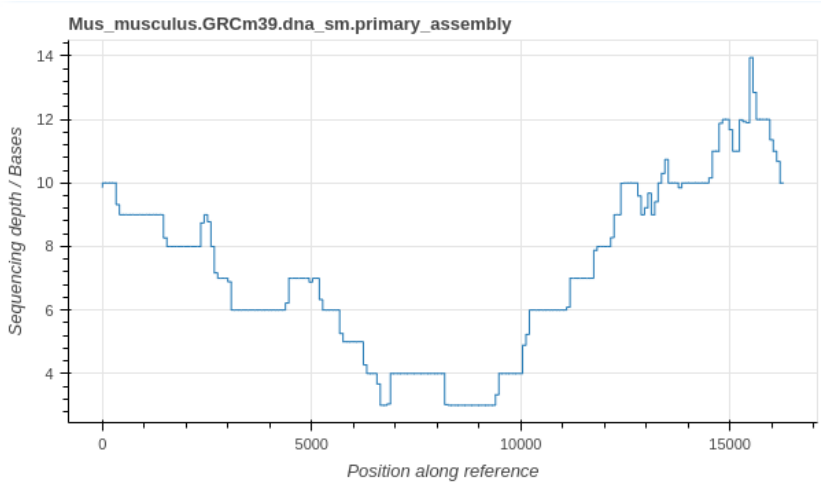
Read depth is an important metric to examine for assessing the efficacy of a sequencing method, because it determines the accuracy of the sequencing run. When sequencing a

heterogenous population and sequencing for epigenetics, read depth also determines the ability of the sequencing run to detect rare genetic and epigenetic variants. Thus, the read depths generated from *MA, MB, MC*, and *MD* were analyzed by graphing the number of sequencing reads that covered each base of the mitogenome during alignment, with the X axis representing position on the mitogenome and the Y axis representing read depth (**Fig. 8**).

MA


Mus_musculus.GRCm39.dna_sm.primary_assembly
Sequencing depth / Bases
160
150
140
130
0
5000
10000
15000
Position along reference

MB


Mus_musculus.GRCm39.dna_sm.primary_assembly
Sequencing depth / Bases
14
12
10
8
6
4
0
5000
10000
15000
Position along reference

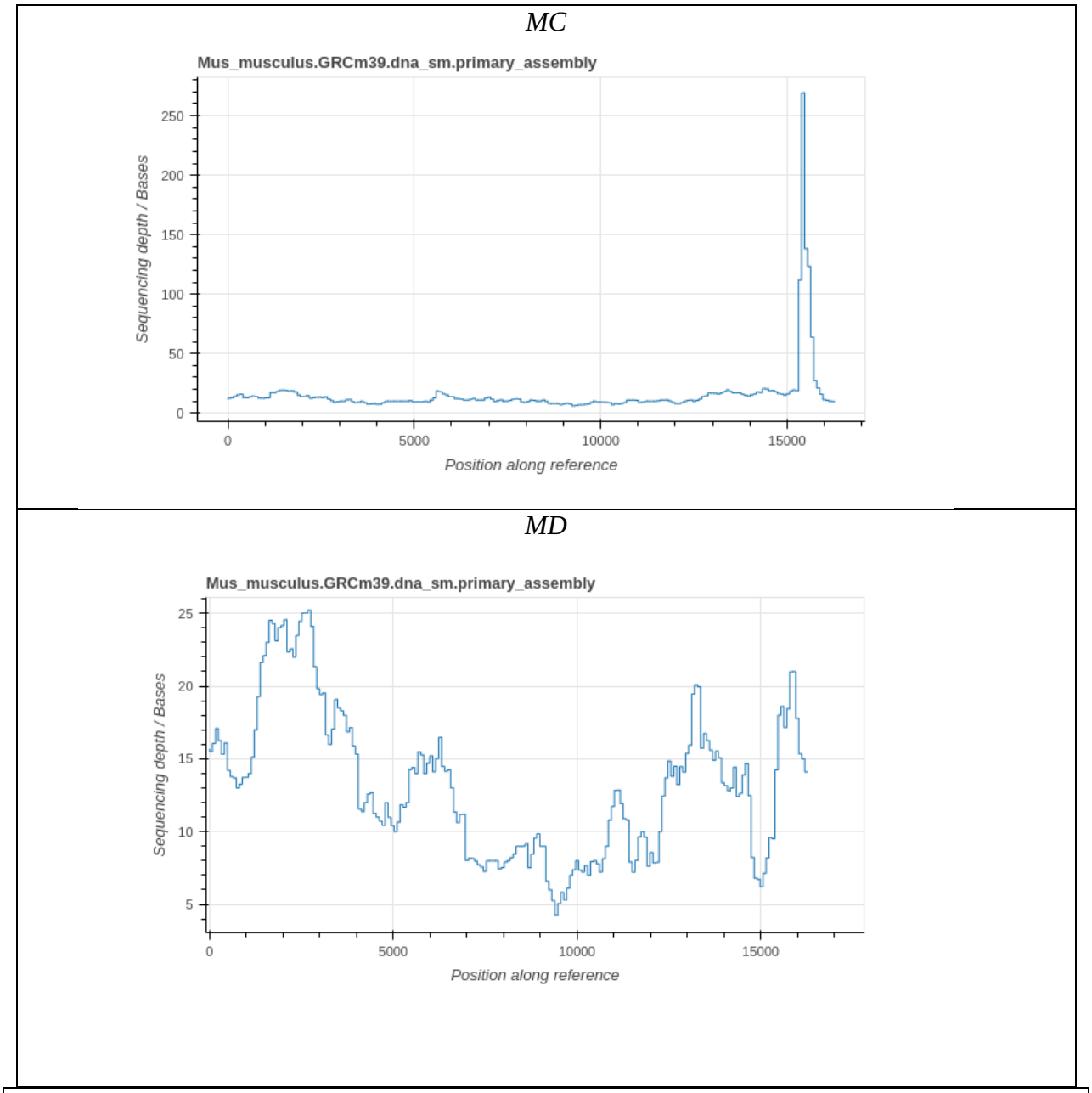


Figure 8. Read depth over mouse mitochondria genome. Graph of the read depth for *MA*, *MB*, *MC*, and *MD*.

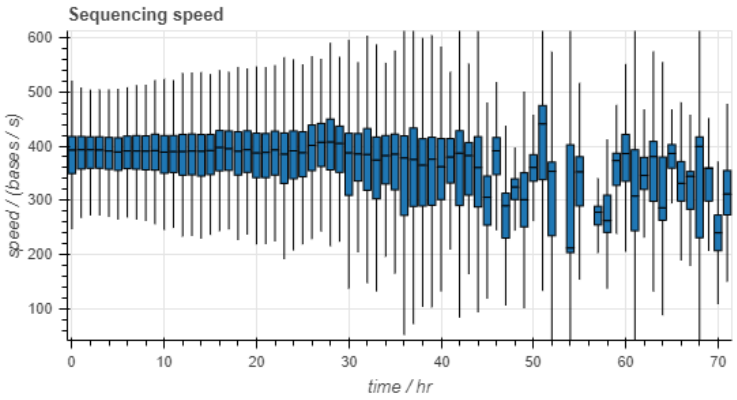
Our results show that *MA* had by far the highest read depth across the mitochondria (**Fig. 8**) and the second highest mitochondrial read count, after *MB* (**Table 2**). *MB* generated the worst read depth and only 2x read depth over parts of the center of the mitochondrial genome (**Fig. 8).** Despite having the second longest average mitochondrial read length, it only generated 22 mitochondrial reads, which is why the read depth is so low (**Table 2**). Notably, the read depth for

MC is the most consistent, around 10x over most of the mitochondria, with a large spike above 250x over parts of the D-loop (between mtDNA bases 15320 and 15720) (**Fig. 8)**. Finally, *MD* has an extremely variable read depth, higher on average than *MB* and *MC* but with a few areas of low coverage (**Fig. 8)**. These results suggest that mtDNA isolation with *MA* (no mtDNA concentrating steps) produces the best read depth when using the Oxford Nanopore Rapid Sequencing Kit (SQK-RAD004) with Oxford Nanopore R9.4 flow cells, whereas *MD* (mitochondrial isolation and plasmid-preference DNA extraction) produces the best results of the mtDNA concentration methods.

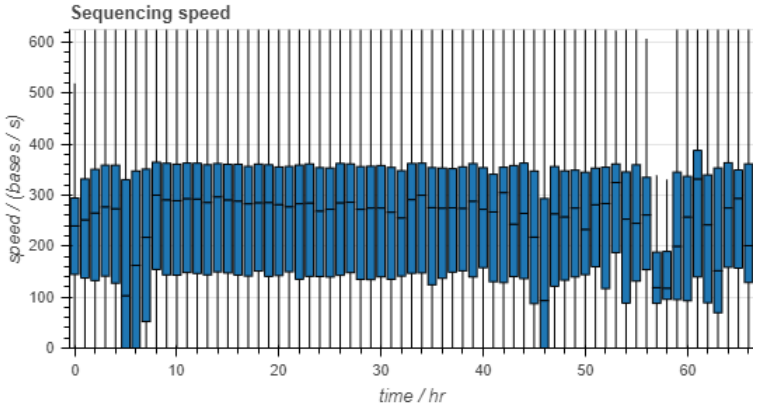
Flow cell degradation rate

To investigate if pore stability was a factor in read count generation, we assessed the performance over time of the flow cells for *MA, MC* and *MD*. The flow cell for *MB* could not be analyzed due to a technical error resulting in the loss of flow cell data for the first portion of the run. The performance of the flow cells used to sequence *MA, MC* and *MD* was assessed using two metrics. The first metric was the change in the sequencing speed of each flow cell over the course of the sequencing operation, which was quantified by measuring the number of bases sequenced per second at hourly intervals (**Fig. 9**). We found that *MA* had the highest average sequencing speed of about 400 bases per second before hour 40, *MC* had an average sequencing speed of about 275 before hour 40, and *MD* has an average of about 300 bases per second before hour 40 (**Fig. 9**). These results suggests that *MA*'s higher total read depth compared to *MD* **(Table 1)** might be due to the amount of sequencing that occurred before channel collapse.

MA



MC



MD

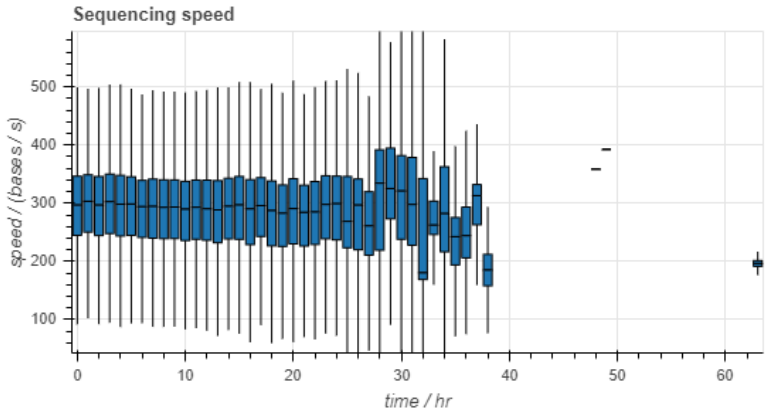
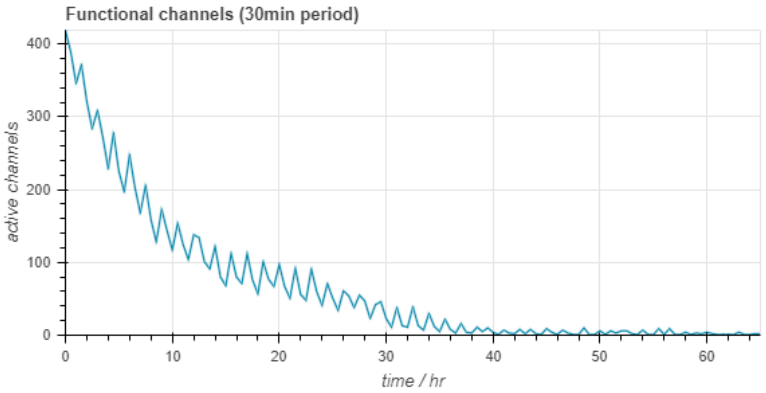


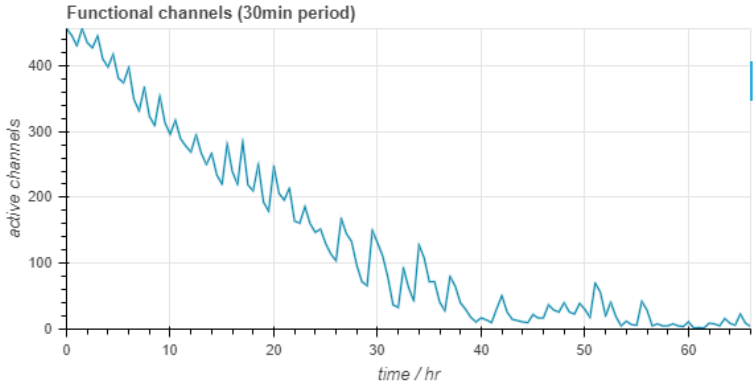
Figure 9. Sequencing speed of flow cell over the course of the sequencing operation. The sequencing speed, bases sequenced per second, for *MA, MC,* and *MD* are presented in hourly increments over the course of the sequencing run.

The second metric assessed was the degradation of the flow cells over the course of the sequencing operation, which was quantified by measuring the number of functional sequencing channels remaining in each flow cell at hourly intervals (**Fig. 10**). We found that *MC*'s and *MD*'s flow cell pores collapsed more slowly than *MA's* (**Fig. 10**). *MC* and *MD* dropped below 200 functioning pore channels at hour 19 and 18 of the run and below 100 functioning channels at hour 29 and 24, respectively (**Fig. 10**). In contrast, *MA* dropped below 200 functioning channels at hour 8, and below 100 functioning channels at hour 12 (**Fig. 10**). *MA* and *MC* retained functional channels for longer than *MD*, with a few channels remaining until the end of the experiment, while *MD* lost all active channels by hour 40 (**Fig. 10**). However, *MA*'s continued high sequencing speed after hour 40, which remained between 200 and 400 bases per second whereas *MC*'s sequencing speed after hour 40 fell to between 100 and 300 bases per second (**Fig. 9**), suggests that despite *MA*'s faster channel collapse, those pores which did remain were able to sequence effectively. These results suggest that the early flow cell collapse in *MC* and *MD* (**Fig. 10**) is associated with an early cessation of sequencing (**Fig. 9**), which impacts the final read depth (**Fig. 8**) and perhaps a lower mitochondrial read count (**Table 2**).

MA



MC



MD

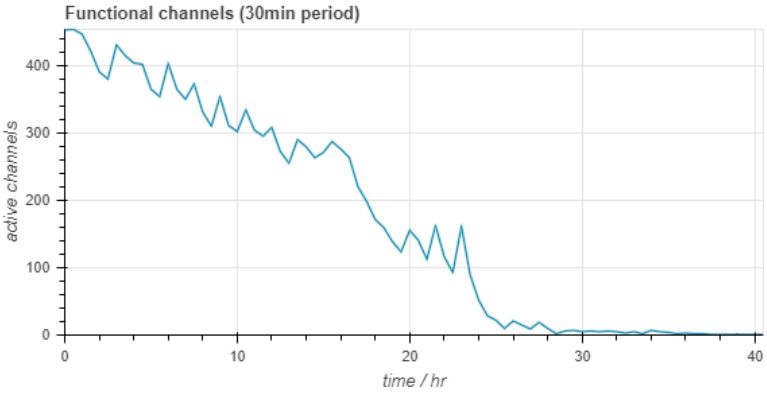


Figure 10. Degradation of the flow cells over the course of the sequencing operation. The number of remaining functioning flow cell channels at each 30 minute increment over the course of the sequencing run for *MA* and *MD*.

Mitogenome methylation detection of different DNA isolation methods

To investigate the role of p107 in potentially regulating mtDNA methylation, p107KO and p107WT were compared using EPI2ME-Labs (Oxford Nanopore, 2023) and the Bioconductor dispersion shrinkage for sequencing data (DSS) analysis (Park and Wu, 2016) programs. *MD* was selected for p107WT because it is the only wildtype sample. *MB* was chosen for the p107KO sample, as it had the most similar read depth distribution to *MD* (**Fig. 8**) and thus introduced the lowest amount of potential bias into the differential methylation analysis. It is important that the two methods being compared possess similar read depths because in Nanopore samples there is a correlation between read depth and the level of methylation detected in a sample if the read depth is below 50 (Lüth, et. al., 2021). The result of this correlation is that if two genomes with identical methylation levels were sequenced, one at a high read depth and one at a low read depth, the one with the high read depth may be incorrectly identified as also having higher methylation levels (Lüth, et. al., 2021).

MD, the only p107WT sample, had a read depth that generally was between 5 and 25 times coverage over the mitochondrial genome (**Fig. 8**). Comparing *MD* to *MA*, which had a read depth of over 130 for most of the mitochondrial genome, or to *MC*, which had a read depth of over 250 in the D-loop (**Fig. 8**), would have resulted in those samples having falsely inflated methylation levels relative to *MD* relative to the levels they actually possess. In contrast, *MB*'s more similar read depth distribution to *MD* means that the detected methylation levels of *MB* and *MC* relative to each other will remain accurate.

Table 5. differentially methylated sites in a p107WT and a p107KO sample. This table displays the genes that exhibited differential methylation, the positions along the mitochondrial genome where differential methylation was observed, the mean methylation frequency for each differentially methylated site in the p107KO sample, the mean methylation frequency for each differentially methylated site in the p107WT sample, the difference in methylation levels for each site between the samples, and the p-value for differential methylation at each site.

Gene	Position	p107KO Methylation	p107WT Methylation	Methylation Difference	p-value
mtATP6	8561	0.20658675	0.09569184	0.11089490	0.1313734
mtATP6	8562	0.20658675	0.09569184	0.11089490	0.2169332
ND4	11459	0.09947919	0.19910718	-0.09962799	0.2575922
ND4	11460	0.09947919	0.19910718	-0.09962799	0.2474921

The methylation alignment data generated by Epi2ME-Labs was input into DSS to identify any genes which had significantly different methylation frequencies between *MB* (p107KO) and *MD* (p107WT). When the methylation levels across *MB* and *MD* were compared with DSS using a p-value cutoff of 0.5, four sites with significant differences in methylation were found, three of which had p-values below 0.25 (**Table 5**). The differentially methylated sites consisted of two adjacent positions for each of mtATP6 and ND4. Because the mouse mtATP6 gene is found between positions 7,927 and 8,607 of the mouse mitogenome (Cunningham et. al., 2022) and the mouse ND4 gene is found between positions 10,167 and 11,544 of the mouse mitogenome (Cunningham et. al., 2022), both differentially methylated regions are located near the 3' end of their respective gene.

Both mtATP6 methylation sites exhibited a difference in mean methylation frequency across all analyzed reads of 0.11089490 more methylation in the p107KO sample, with p-values of 0.1313734 and 0.2169332 for mitogenome positions 8561 and 8562, respectively (**Table 5**). For the adjacent differentially methylated bases in the ND4 gene, both exhibited a difference in mean methylation frequency across all analyzed reads of 0.09962799, with more methylation in the

p107WT sample (**Table 5**). The p-values are 0.2575922 and 0.2474921 for mitogenome positions 11459 and 11460, respectively (**Table 5**).

Additional analyses were also run comparing the mtDNA methylation levels of the p107KO *MA* and *MC* to *MD* (see **Appendix 1** and **Appendix 2**). Though their higher read depths relative to *MD* make these analyses less reliable comparators, their higher read depths relative to *MB* will produce results with a superior p-value. When *MA* was compared to *MD* with a p-value cutoff of 0.05, 53 sites distributed across the ND2, ND4, ND5, and tRNA-Trp genes in *MA* were found to have methylation frequencies of between 0.12 and 0.15 lower than in *MD* with p-values below 0.05 (**Appendix 1**). The ND2, ND4 and ND5 genes encode for components of OXPHOS complexes, and tRNA-trp is a transfer RNA. When *MC* was compared to *MD* with a p-value cutoff of 0.05, 119 sites distributed across the ND1, ND2, ND4, ND5, CO1, mtATP6, tRNA-Trp, and tRNA-Gln genes were found to have frequencies in *MC* between 0.12 and 0.15 lower than in *MD* with p-values below 0.05. (**Appendix 2**) The ND1, ND2, ND4, ND5, CO1, and mtATP6 genes encode for components of OXPHOS complexes, and tRNA-Trp and tRNA-Gln are transfer RNAs. All three p107KO samples displayed reduced methylation relative to *MD* on the ND4 gene, though not at the same site. *MB* displayed reduced ND4 methylation at bases 11459 and 11460 (**Table 5**), while *MA* and *MC* both displayed reduced ND4 methylation at base 11420.

The p107KO *MA* and *MC* samples both displayed reduced levels of ND2, tRNA-trp, and ND5 methylation relative to the p107WT *MD* (**Appendix 1** and **Appendix 2**). Both samples showed reduced ND2 methylation between bases 4920 and 2936. Both samples showed reduced tRNA-trp methylation at bases 4939, 4946, 4947, 4950, 4953, 4958, 4959, 4965, and 4968. Both samples showed reduced ND5 methylation at bases 12778, 12782, 12786, 12787, 12788, 12790, 12857, and 12860. The *MB* and *MC* samples both displayed differential methylation in the

mtATP6 gene relative to *MD*, but in different directions. With *MB* displayed increased methylation in the gene at sites 8561 and 8562, and *MC* displaying decreased methylation in the gene. Taken together, the results of all three differential methylation analyses suggest that the mitochondrial methylome is substantively altered in the absence of p107, primarily through reduced methylation of ND4 and other OXPHOS components, and therefore that p107 regulates mtDNA methylation.

Chapter 4: Discussion

When comparing the sequencing efficiency of the different methods used to obtain mtDNA, we found that *MA* (without direct mtDNA isolation) generated the best final read depth (**Fig. 8**). *ME* (mitochondrial isolation followed by plasmid-preference DNA extraction, and adjustment for cell culture mtDNA content) generated the best percentage of reads with passing Q-scores which aligned to the mitochondria (**Table 2**). *MC* had the highest percentage of passing mitochondrial reads after *ME*. When the p107WT (*MD)* and the p107KO (*MB)* were compared to identify differentially methylated sites in the mitochondrial genome, mtATP6 was found to exhibit higher methylation levels in the p107KO at adjacent bases, whereas ND4 was found to exhibit higher methylation levels in the p107WT at adjacent bases. ND4 was also found to exhibit higher methylation levels in the p107WT when the p107KO *MA* and *MC* were compared to *MD*.

Nanopore sequencing, compared to other widely used methods such as bisulfite sequencing, is better at sequencing over repetitive regions (Heydari et. al., 2019, Kinkar et. al., 2021) and correlating polymorphisms (Huang et. al., 2023) into haplotypes. These capabilities are a result of the longer reads generated by Nanopore sequencing (Quail, Swerdlow, and Turner, 2009, Amarasinghe et. al., 2020). Moreover, Nanopore's signal-based methylation detection method (Oxford Nanopore 2023) provides additional benefits, allowing data to be analyzed for multiple types of epigenetic modification (Rand et. al., 2017, Sood, Viner, and Hoffman, 2019, Liu et. al., 2015, Kot et. al., 2020). Bisulfite sequencing only detects methylation, whereas Nanopore-generated data can be analyzed for any type of epigenetic modification that has had a sufficiently accurate detection algorithm written for it (Rand et. al., 2017, Sood, Viner, and Hoffman, 2019, Liu et. al., 2015, Kot et. al., 2020). This is particularly important because new types of epigenetic base modification beyond acetylation and 5mC methylation continue to be discovered (Luo et. al.,

2015, Maejima and Sadoshima, 2014, Kumar, Chinnusam and Mohapatra, 2018). Importantly, as base modification detection algorithms can be applied to existing sequencing data from previous Nanopore experiments, any modification can be identified and investigated using existing data without the need to collect new samples and perform new sequencing experiments. Similarly, if the accuracy of a base modification detection algorithm is improved, the new, more accurate version of that algorithm can be reapplied to an older experiment, allowing for an improved confidence in that experiment's results without needing to perform the experiment again. Another advantage Nanopore sequencing has over both restriction enzyme methylation detection and bisulfide sequencing is that Nanopore sequencing can be performed in the field (Runtuwene et. al., 2019). The ability to sequence in the field allows for fresh samples to be used that have not been subjected to storage or transportation and thus have had less time for degradation of the sample to occur (Pramanik, Sarma, and Murty, 2006).

The regulation of OXPHOS by p107 makes it a potential regulator of DNA epigenetic modifications by modulating the supply of the metabolites necessary for DNA base modification. The synthesis of SAM is dependent on the rate of cellular metabolism and whether OXPHOS or glycolysis is dominant within a cell (Ouyang et. al., 2020, Shi and Tu, 2016). Thus, when p107 influences the cell's metabolic rate it will impact the availability of SAM and thus the cell's ability to carry out epigenetic modifications.

On their own plasmid-preference DNA extraction or isolation of the mitochondria before DNA extraction have both been shown to increase mtDNA concentration and mitochondrial read depth (Quispe-Tintaya et. al., 2013). However, when both approaches were used in the same method to extract mtDNA (*MD*) it resulted in a worse read depth than *MA* where total DNA was used, (**Fig. 8**). This result can be partially explained when the different components that contribute

to read depth are examined. Read depth is a function of both read count and read length. In this case, 1.45% of reads from *MD* with passing Q-scores aligned to the mitochondria whereas only 0.21% of passing *MA* reads aligned to the mitochondria (**Table 2**). Thus, *MD* would be expected to generate 6.9% as many mitochondrial reads as *MA* if reads in both samples were the same length and were being sequenced at the maximum rate. However, this was not the case because *MA* sequenced at a higher speed and for longer than *MD* (**Fig. 9 & Fig. 10**). Thus, the total number of reads generated for each method had a larger influence on the number of mitochondrial reads generated, resulting in *MA* generating more mitochondrial reads with passing Q-scores than *MD* (**Table 1**). In addition, the higher number of mitochondrial reads generated by *MA* relative to *MD* might be due to *MA* using the recommended input DNA concentration for the Oxford Nanopore SQK-RAD004 that is 400µg, compared to the 99µg of input DNA that was used for *MD* (**Table 1**). The low quantity of input *MD* DNA was a result of the limited concentration of total DNA that was obtained using this method.

Another factor that might also have impacted the read count for *MD* is that it is the only sample generated from p107WT cells, while the other methods used p107KO cells. In this case, p107WT cells have lower mtDNA content than p107KO cells (Bhattacharya, Shah, and Oresajo, 2021), and when this was accounted for with *ME*, the outcome was 3.18 times as many mitochondrial reads with passing Q-scores as in *MA* (**Table 1**).

Our results suggest that the inclusion of plasmid-preference DNA extraction can increase mtDNA concentration over what can be achieved with mitochondrial isolation alone. As mitochondrial isolation was also used in *MB*, its results can be analyzed to provide insight into the potential cause of the lower read depth for *MD*. *MB* generated the lowest number of passed mitochondrial reads, generating only 22 in comparison to the 368 passed mitochondrial reads

generated by *MA* (**Table 2**), which had no mtDNA concentration techniques applied to it. Mitochondrial isolation has been shown to increase mtDNA concentration and read depth (Quispe-Tintaya et. al., 2013), thus it is possible that the poor performance for *MB* is a result of a shortcoming with the specific mitochondrial organellar isolation technique used, which used physical agitation with a Dounce homogenizer to break down the cellular membrane. This possibility is further supported by evidence that mitochondrial isolation using centrifuge-based methods without douncing can increase mtDNA concentration in a sample (Quispe-Tintaya et. al., 2013). If this shortcoming in mitochondrial extraction methodology is the cause of *MB*'s poor read count, then the same shortcoming would also reduce *MD*'s read count. *MD* generated a higher read depth than *MB* (**Fig. 8**). Because these two methods only differed in the addition of plasmid-preference DNA extraction in *MD*, the read depth difference suggests that the combination of plasmid-preference DNA extraction with mitochondrial isolation is superior at mtDNA concentration than mitochondrial isolation alone.

In addition to read count, read length also impacted the difference in mitogenome read depth between *MD* and *MA*. The samples had similar Q-score passing mitochondrial read counts, with 234 for *MD* and 368 for *MA* (**Table 1**), but the passed mitochondrial reads for *MA* were on average 11.23 times longer than for *MD*. This would potentially contribute to *MA* having the superior read depth. Because physical agitation is known to fragment DNA (Li et. al., 2017), the additional processing steps involved in generating *MD*, where the mitochondria is isolated before DNA extraction, and DNA isolation is performed using a spin column kit which preferentially extracts circularized DNA (Quispe-Tintaya et. al., 2013), might increase fragmentation sufficiently to impact the final read depth.

In addition to common bacteria such as *Escherichia* and *Salmonella*, the list of non-mouse organisms whose DNA was found in the experimental samples included *Muviruses*, and gamma retroviruses (**Table 3**). This suggests contamination of the cell cultures but could also be a result of DNA from gamma retroviruses that integrated into the genome of the mouse population from which C2C12 cells were derived. Retrovirus integration into the mouse genome is a well-known phenomenon (Garen 2016). The fact that these reads were not present in the same quantities in *MC* and *MD* (**Table 3**), which contained fewer autosomal mouse reads (**Table 2**), supports this possibility. Also, the two most common gamma retroviruses in the sample, Mus musculus mobilized endogenous polytropic provirus and Murine type C retrovirus, both successfully aligned to the genome of C3H, the mouse strain from which the C2C12 cell line was derived (Yaffe and Saxel, 1977).

Though the level of contamination that was present allowed sequencing to a sufficient read depth for analysis (**Fig. 8**, **Table 5**, **Appendix 1** and **Appendix 2**) there is utility in reducing further its presence. Multiple techniques exist, which could be employed to reduce sample contamination. Sterile, automated systems can be used to perform spin column based operations (Qiagen, 2024), thereby reducing the exposure of the sample to open air and the possibility of contaminants being introduced during manual handling. Additionally, spin column methods exist which can be used for mitochondrial isolation as an alternative to Dounce homogenization (Thermo Scientific, 2023), which would remove the possibility of contaminations being introduced on the homogenizer. Alternatively, the use of disposable homogenizers or autoclaving of a reusable homogenizer between experiments could also reduce the exposure to contaminants.

Breast, cervical, colorectal, gastric, hepato-cellular, lung ovarian, pancreatic, and prostate cancers have all been shown to be subject to changes in mtDNA methylation in some individuals

which can make the cancers resistant to some anti-cancer drugs (Romero-Garcia, Prado-Garcia, Carlos-Reyes, 2020). Increasing mtDNA concentration can help to screen patient cancer samples for the methylation markers which induce drug resistance by allowing for the multiplexing of cancer DNA samples to make the epigenetic sequencing of these samples more efficient.

Inorganic contaminants are a potential driver of poor sequencing performance, because non-DNA contaminants are known to contribute to pore collapse (Chou et. al., 2020). *MD* stopped sequencing before *MA* and had its flow cell degrade faster than *MA*'s (**Fig. 9**) (**Fig. 10**). Because *MD* contained an additional mitochondrial isolation step not found in *MA*, it is possible that its faster flow cell channel collapse may have been caused by the additional processing steps involved in mitochondrial isolation. This is supported by *MC*, which also used plasmid-preference DNA extraction but did not possess the additional mitochondrial removal step, retained functional channels and continued to sequence until the end of the sequencing run (**Fig. 9** & **Fig. 10**)

The ND4 gene displays decreased methylation in all three p107KO samples, *MA, MB,* and *MC* compared to the p107WT *MA* (**Table** 5, **Appendix 1** and **Appendix 2**). ND4 is a component of OXPHOS complex I (Alharbi et. al., 2019). Loss of function mutations in ND4 have been shown to inhibit OXPHOS activity (Xu et. al., 2017), thus its methylation in p107WT cells is consistent with a p107 role as a suppressor of OXPHOS (Bhattacharya et. al., 2021). Loss of function mutations of the ND4 gene have been found in human patients with multiple sclerosis, an autoimmune disease of the central nervous system (Alharbi et. al., 2019). p107 has been shown to play a role in regulating neural stem cell populations by preventing apoptosis and supressing cell growth (Robanus-Maandag et. al., 1998). The ND2, ND5, and tRNA-Trp genes also displayed reduced methylation in *MA* and *MC* relative to *MD* (**Appendix 1** and **Appendix 2**), though they do not display any difference in methylation levels between *MB* and *MD* (**Table 5**). *MC* also

displayed decreased methylation of the tRNA-Gln gene relative to *MD* and was the only knockout sample to do so. These results are also in line with our hypothesis, because ND2 and ND5 are components of OXPHOS complex I (Alharbi et. al., 2019). tRNA-Trp and tRNA-Gln encode tRNAs which are necessary for proper OXPHOS function. Mitochondrial mutations which reduce tRNA-Trp expression are correlated with reduced OXPHOS function in human patients with Combined OXPHOS system deficiencies (Smits et. al., 2009). Mutations which impede the amino acetylation of tRNA-Gln, and thus its ability to participate in protein translation, are associated with reduced OXPHOS complex biosynthesis and a resulting reduction in OXPHOS function (Friederich et. al., 2018).

The observed increase in methylation of the mtATP6 gene, which is a subunit of the ETC, in *MB* (p107KO) is surprising as it would imply a decrease in mtATP6 gene expression and OXPHOS activity (Moreno-Loshuertos et. al., 2023), despite p107KO displaying increased OXPHOS (Bhattacharya et. al., 2021). However, loss of function mutations in mtATP6 are known to be tumorigenic (Moreno-Loshuertos et. al., 2023). Hence, it is possible that the suppression of mtATP6 in the absence of p107 is a result of p107's role as a tumor suppressor (Dannenberg et. al., 2004) that is unrelated to OXPHOS control. It is also, however, possible that the anomalous result is a false positive and that mtATP6 is not in fact methylated in the absence of p107. This is supported by the low p-values, 0.13 and 0.22, for mtATP6 (**Table 5**), and by the decreased methylation of mtATP6 in *MC* relative to *MD*, which was observed with a p-value of around $7e^{-11}$ at most sites (**Appendix 2**), making this the more likely possibility. This could be checked by performing additional sequencing runs of *MB* and pooling the resulting reads, and doing the same for *MD*, in order to improve the read depth and the resulting accuracy of the read base calling (Castro-Wallace et. al., 2017) (Liu et. al., 2021).

MA and *MC* have higher read depths across the mitogenome than *MD* (**Fig. 8**), and therefore have the potential to introduce bias into the differential methylation calculations (Lüth, et. al., 2021). However, the bias associated with higher read depths is for an increase in the amount of methylation detected at any given site, and so any sites which display increased methylation in the higher read depth samples may be an artifact of bias. *MA* and *MC*, however, did not display any sites with increased methylation relative to *MD*, only sites with decreased methylation relative to *MD* (**Appendix 1** and **Appendix 2**). Thus, if the results were altered by the read depth bias, it would tend to weaken or eliminate any results with decreased methylation relative to *MD*. Hence, any which remain have a high confidence of being real effects.

One major shortcoming of this thesis is the lack of experimental replicates for each of the methods, as *MA, MB, MC,* and *MD* had only one replicate each. This was done to investigate the utility of nanopore sequencing of mtDNA using several approaches to its isolation. However, error arising from chance events in the samples used could have contributed to the differences in results between the different methods used for mtDNA extraction. Repeating the experiment with more replicates would enable us to investigate whether the observed differences in mtDNA methylation patterns between p107WT and p107KO myoblasts are a true difference or a result of sampling error.

The addition of more replicates would improve the results of the comparison of mtDNA concentration methods. We have theorized that *MB* and *MC*'s low read depth relative to *MA* could be result of loss of DNA, the addition of contaminants, or damage to the DNA occurring as a result of the mitochondrial isolation employed in *MB* and the plasmid-preference DNA extraction employed in *MC*. It is an important factor to note, however, that while *MB* and *MC* were derived from the same p107KO culture, *MA* was derived from a different p107KO culture. It is possible

that a random event during the CRISPR/Cas9 process used to generate these knockouts resulted in one strain possessing higher mtDNA content than another, which would skew those results, and performing additional replicates using multiple wildtype, scrambled, and knockout samples for each method would eliminate this possibility and allow for a more direct comparison between methods. Similarly, additional replicates would also eliminate the need for *ME* and its artificial adjustment for differences in mtDNA content between cultures, because the same cultures could be compared between all mtDNA concentration methods, and any differences in mtDNA content between cultures could be directly compared.

In addition to their utility in studies on mitochondrial function, methods to concentrate and extract mtDNA also have ecological applications. The ability to collect mtDNA data from several individuals can facilitate surveys of the genetic diversity of animal populations (Wanner et. al., 2021). This is especially valuable because it allows for non-invasive sampling of rare populations, including endangered species because the high per-cell copy number of mtDNA allows it to be acquired in sufficient quantities to be sequenced from fecal samples, which can be acquired without directly interacting with individual animals (Wanner et. al., 2021).

Performing mtDNA surveys using Nanopore sequencing has the added benefit of being able to incorporate epigenetic data, which is important to provide information on the distribution of individuals at different life stages within that population (Sirard 2019). Some species including humans display changes to mtDNA methylation patterns during the aging process (Sirard 2019). These markers can be used to guess the age of the individual which left a specific stool sample, and thus can be used to estimate the age distribution of a population if enough samples are taken. Humans have also been shown to exhibit changes in mtDNA methylation patterns in response to exposure to environmental pollutants (Sharma, Pasala, and Prakash, 2019). This could be applied

to other mammals, where fecal mtDNA samples could be compared over time or between locations with different pollutant exposure to identify the degree to which a species of interest's cellular activity is affected by the pollutants around it, in order to gain an understanding of its vulnerability to environmental changes.

There are additional steps that could be taken to improve the statistical confidence in the results found in this thesis. Greater confidence in the results of the comparative analysis of extraction methods *MA*, *MB*, *MC*, and *MD* could be achieved by increasing the number of replicates and performing each mtDNA extraction method again on the same source strain. This would control for any potential technical errors in performing the extractions. It will help determine whether the poor performance of *MB* and *MC* compared to their demonstrated potential for mtDNA concentration (Quispe-Tintaya et. al., 2013) was a result of the human error during the extraction process.

Because Nanopore sequencing allows for the pooling of results from different sequencing runs, *MB, MC,* and *MD* could be sequenced additional times to bring their read depths above 50, the threshold at which read depth differences do not introduce methylation detection bias (Lüth, et. al., 2021). This would allow improvement of the p-value for the differential methylation analysis of *MB* against *MD*. It will also allow us to investigate whether any genes with lower methylation levels in *MA* or *MC* relative to *MD* were missed during the initial analysis due to their methylation levels being inflated by read depth induced bias.

In addition, the statistical confidence of both the extraction method and differential methylation analyses could be improved by introducing more biological replicates. If each extraction method was performed on multiple p107KO and p107WT C2C12 strains, it would control for possible differences in mtDNA content in the absence of p107, and for any off-target

effects of the CRISPR process used to generate the KO cells, which could impact mtDNA content. It would also allow us to ensure that the WT strain used for *MD* is in fact representative of the mitogenomes of WT C2C12 cells. Additionally, the usage of a cultured myoblast cell line derived from a mouse strain other than C3H would allow us to control for and investigate any differences in p107's effect on mtDNA methylation between different mouse strains.

There are also steps that can be taken to expand on the results in future experiments. Though cell cultures are commonly used to study biological phenomena, they do not perfectly replicate the behaviour of cells in vivo (Edmonson, Broglie, Adcock and Yang, 2014). This can be corrected by sequencing satellite cells and myoblasts from tissue samples (Munchel et. al., 2015). The cells can be isolated from tissue samples using Fluorescence-activated cell sorting (Gromova, Tierney, and Sacco1, 2015) (Pakula, Spinazolla, and Gussoni, 2019). This will allow us to better understand the effects of p107 on mtDNA methylation in living organisms.

In conclusion, the results of this paper demonstrate that by combining mitochondrial isolation and plasmid-preference DNA extraction, mtDNA can be isolated more effectively than with either method alone, and that p107 regulates mtDNA methylation at OXPHOS genes.

References:

1. Alharbi, M.A., Al-Kafaji, G., Khalaf, N.B., Messaoudi, S.A., Taha, S., Daif, A., & Bakhiet, M. (2019). Four novel mutations in the mitochondrial ND4 gene of complex I in patients with multiple sclerosis. Biomedical Reports, 11, 257-268. https://doi.org/10.3892/br.2019.1250
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10. doi: 10.1016/S0022-2836(05)80360-2. PMID: 2231712.
3. Amarasinghe, S.L., Su, S., Dong, X. et al. Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21, 30 (2020). https://doi.org/10.1186/s13059-020-1935-5
4. Araki, K., Kawauchi, K., Sugimoto, W. et al. Mitochondrial protein E2F3d, a distinctive E2F3 product, mediates hypoxia-induced mitophagy in cancer cells. Commun Biol 2, 3 (2019). https://doi.org/10.1038/s42003-018-0246-9
5. Basu U, Bostwick AM, Das K, Dittenhafer-Reed KE, Patel SS. Structure, mechanism, and regulation of mitochondrial DNA transcription initiation. J Biol Chem. 2020 Dec 25;295(52):18406-18425. doi: 10.1074/jbc.REV120.011202. Epub 2020 Oct 30. PMID: 33127643; PMCID: PMC7939475.
6. Bayona-Bafaluy MP, Acín-Pérez R, Mullikin JC, Park JS, Moreno-Loshuertos R, Hu P, Pérez-Martos A, Fernández-Silva P, Bai Y, Enríquez JA. Revisiting the mouse mitochondrial DNA sequence. Nucleic Acids Res. 2003 Sep 15;31(18):5349-55. doi: 10.1093/nar/gkg739. PMID: 12954771; PMCID: PMC203322.

7. Bellizzi D, D'Aquila P, Giordano M, Montesanto A, Passarino G. Global DNA methylation levels are modulated by mitochondrial DNA variants. Epigenomics. 2012 Feb;4(1):17-27. doi: 10.2217/epi.11.109. PMID: 22332655.

8. Bhattacharya D, Scimè A. Mitochondrial Function in Muscle Stem Cell Fates. Front Cell Dev Biol. 2020 Jun 16;8:480. doi: 10.3389/fcell.2020.00480. PMID: 32612995; PMCID: PMC7308489.

9. Bhattacharya D. , Ydfors M. , Hughes M. C. , Norrbom J. , Perry C. G. R. , Scimè A. Decreased transcriptional co-repressor p107 is associated with exercise-induced mitochondrial biogenesis in human skeletal muscle, Physiol Rep, 5 (5), 2017,e13155, doi: 10.14814/phy2.13155

10. Bhattacharya, D. (2017). Metabolic Regulation by p107 (Rbl1) Influences Muscle Stem Cell Fate Decisions [Doctoral dissertation, York University]. York University Electronic Theses and Dissertations. http://hdl.handle.net/10315/38245

11. Bhattacharya, D., Shah, V., Oresajo, O. et al. p107 mediated mitochondrial function controls muscle stem cell proliferative fates. Nat Commun 12, 5977 (2021). https://doi.org/10.1038/s41467-021-26176-0

12. Blanch M, Mosquera JL, Ansoleaga B, Ferrer I, Barrachina M. Altered Mitochondrial DNA Methylation Pattern in Alzheimer Disease-Related Pathology and in Parkinson Disease. Am J Pathol. 2016 Feb;186(2):385-97. doi: 10.1016/j.ajpath.2015.10.004. PMID: 26776077.

13. Bogenhagen DF, Applegate EF, Yoza BK. Identification of a promoter for transcription of the heavy strand of human mtDNA: in vitro transcription and deletion mutagenesis. Cell. 1984 Apr;36(4):1105-13. doi: 10.1016/0092-8674(84)90061-8. PMID: 6323020.

14. Bradley-Whitman MA, Lovell MA. Epigenetic changes in the progression of Alzhei’er's disease. Mech Ageing Dev. 2013 Oct;134(10):486-95. doi: 10.1016/j.mad.2013.08.005. Epub 2013 Sep 3. PMID: 24012631; PMCID: PMC3857018.

15. Brandvain Y, Wade MJ. The functional transfer of genes from the mitochondria to the nucleus: the effects of selection, mutation, population size and rate of self-fertilization. Genetics. 2009 Aug;182(4):1129-39. doi: 10.1534/genetics.108.100024. Epub 2009 May 17. PMID: 19448273; PMCID: PMC2728854.

16. Castro-Wallace SL, Chiu CY, John KK, Stahl SE, Rubins KH, McIntyre ABR, Dworkin JP, Lupisella ML, Smith DJ, Botkin DJ, Stephenson TA, Juul S, Turner DJ, Izquierdo F, Federman S, Stryke D, Somasekar S, Alexander N, Yu G, Mason CE, Burton AS. Nanopore DNA Sequencing and Genome Assembly on the International Space Station. Sci Rep. 2017 Dec 21;7(1):18022. doi: 10.1038/s41598-017-18364-0. PMID: 29269933; PMCID: PMC5740133.

17. Cheng GP, Guo SM, Yin Y, Li YY, He X, Zhou LQ. Aberrant Expression of Mitochondrial SAM Transporter SLC25A26 Impairs Oocyte Maturation and Early Development in Mice. Oxid Med Cell Longev. 2022 Apr 13;2022:1681623. doi: 10.1155/2022/1681623. PMID: 35464759; PMCID: PMC9020962.

18. Chestnut BA, Chang Q, Price A, Lesuisse C, Wong M, Martin LJ. Epigenetic regulation of motor neuron cell death through DNA methylation. J Neurosci. 2011 Nov

16;31(46):16619-36. doi: 10.1523/JNEUROSCI.1639-11.2011. PMID: 22090490; PMCID: PMC3238138.

19. Chou YC, Masih Das P, Monos DS, Drndić M. Lifetime and Stability of Silicon Nitride Nanopores and Nanopore Arrays for Ionic Measurements. ACS Nano. 2020 Jun 23;14(6):6715-6728. doi: 10.1021/acsnano.9b09964. Epub 2020 Apr 27. PMID: 32275381; PMCID: PMC9547353.

20. Cunningham F, et. al. Ensembl 2022. Nucleic Acids Res. 2022, vol. 50(1):D988-D995 PubMed PMID: 34791404. doi:10.1093/nar/gkab1049

21. Dannenberg JH, Schuijff L, Dekker M, van der Valk M, te Riele H. Tissue-specific tumor suppressor activity of retinoblastoma gene homologs p107 and p130. Genes Dev. 2004 Dec 1;18(23):2952-62. doi: 10.1101/gad.322004. PMID: 15574596; PMCID: PMC534655.

22. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. PLoS One. 2021 Oct 1;16(10):e0257521. doi: 10.1371/journal.pone.0257521. PMID: 34597327; PMCID: PMC8486125.

23. D'Souza AR, Minczuk M. Mitochondrial transcription and translation: overview. Essays Biochem. 2018 Jul 20;62(3):309-320. doi: 10.1042/EBC20170102. PMID: 30030363; PMCID: PMC6056719.

24. Dzitoyeva S, Chen H, Manev H. Effect of aging on 5-hydroxymethylcytosine in brain mitochondria. Neurobiol Aging. 2012 Dec;33(12):2881-91. doi: 10.1016/j.neurobiolaging.2012.02.006. Epub 2012 Mar 22. PMID: 22445327; PMCID: PMC3462297.

25. Evans LH, Alamgir AS, Owens N, Weber N, Virtaneva K, Barbian K, Babar A, Malik F, Rosenke K. Mobilization of endogenous retroviruses in mice after infection with an exogenous retrovirus. J Virol. 2009 Mar;83(6):2429-35. doi: 10.1128/JVI.01926-08. Epub 2008 Dec 30. PMID: 19116259; PMCID: PMC2648265.

26. Lopes, AFC. Mitochondrial metabolism and DNA methylation: a review of the interaction between two genomes. Clin Epigenet 12, 182 (2020). https://doi.org/10.1186/s13148-020-00976-5

27. Feng S, Xiong L, Ji Z, Cheng W, Yang H. Correlation between increased ND2 expression and demethylated displacement loop of mtDNA in colorectal cancer. Mol Med Rep. 2012 Jul;6(1):125-30.

28. Flynn R, Washer S, Jeffries AR, Andrayas A, Shireby G, Kumari M, Schalkwyk LC, Mill J, Hannon E. Evaluation of nanopore sequencing for epigenetic epidemiology: a comparison with DNA methylation microarrays. Hum Mol Genet. 2022 Sep 10;31(18):3181-3190. doi: 10.1093/hmg/ddac112. PMID: 35567415; PMCID: PMC9476619.

29. Friederich MW, et. al. Pathogenic variants in glutamyl-tRNAGln amidotransferase subunits cause a lethal mitochondrial cardiomyopathy disorder. Nat Commun. 2018 Oct 3;9(1):4065. doi: 10.1038/s41467-018-06250-w. PMID: 30283131; PMCID: PMC6170436.

30. Garen A. From a retrovirus infection of mice to a long noncoding RNA that induces proto-oncogene transcription and oncogenesis via an epigenetic transcription switch. Signal Transduct Target Ther. 2016 May 13;1:16007. doi: 10.1038/sigtrans.2016.7. PMID: 29263895; PMCID: PMC5661657.

31. Ghosh, Sourav & Singh, Keshav & Sengupta, Shantanu & Scaria, Vinod. (2015). Mitoepigenetics: The different shades of grey. Mitochondrion. 25. 10.1016/j.mito.2015.09.003.

32. Gromova A, Tierney MT, Sacco A. FACS-based Satellite Cell Isolation From Mouse Hind Limb Muscles. Bio Protoc. 2015 Aug 20;5(16):e1558. doi: 10.21769/bioprotoc.1558. PMID: 27668269; PMCID: PMC5034768.

33. Hahn MA, Szabó PE, Pfeifer GP. 5-Hydroxymethylcytosine: a stable or transient DNA modification? Genomics. 2014 Nov;104(5):314-23. doi: 10.1016/j.ygeno.2014.08.015. Epub 2014 Aug 30. PMID: 25181633; PMCID: PMC4252803.

34. Hashimoto K, Kokubun S,Itoi E and Roach HI (2007) Improved Quantification of DNA Methylation Using Methylation-Sensitive Restriction Enzymes and Real-Time PCR, Epigenetics, 2:2, 86-91, DOI: 10.4161/epi.2.2.4203

35. Hayatsu H. Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis--a personal account. Proc Jpn Acad Ser B Phys Biol Sci. 2008;84(8):321-30. doi: 10.2183/pjab.84.321. PMID: 18941305; PMCID: PMC3722019.

36. Heinemeyer, 2005. Retroviridae complete genome (murine type C retrovirus). GenBank: X04150.1. accessed February 19th, 2024. https://www.ncbi.nlm.nih.gov/nuccore/X94150

37. Heydari M, Miclotte G, Van de Peer Y, Fostier J. Illumina error correction near highly repetitive DNA regions improves de novo genome assembly. BMC Bioinformatics. 2019 Jun 3;20(1):298. doi: 10.1186/s12859-019-2906-2. PMID: 31159722; PMCID: PMC6545690.

38. Hoshi T, Heinemann S. Regulation of cell function by methionine oxidation and reduction. J Physiol. 2001 Feb 15;531(Pt 1):1-11. doi: 10.1111/j.1469-7793.2001.0001j.x. PMID: 11179387; PMCID: PMC2278441.

39. Huang N, Xu M, Nie F, Ni P, Xiao CL, Luo F, Wang J, NanoSNP: a progressive and haplotype-aware SNP caller on low-coverage nanopore sequencing data, Bioinformatics, Volume 39, Issue 1, January 2023, btac824, https://doi.org/10.1093/bioinformatics/btac824

40. Illumina 2010 De Novo Assembly Using Illumina Reads, Illumina, Inc., accessed October 10th, 2023 https://www.illumina.com/Documents/products/technotes/technote_denovo_assembly_ecoli.pdf

41. Jin Q, Yuan Z, Xu J. et. al. Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. Nucleic Acids Res. 2002 Oct 15;30(20):4432-41. doi: 10.1093/nar/gkf566. PMID: 12384590; PMCID: PMC137130.

42. Julian LM, Blais A. Transcriptional control of stem cell fate by E2Fs and pocket proteins. Front Genet. 2015 Apr 28;6:161. doi: 10.3389/fgene.2015.00161. PMID: 25972892; PMCID: PMC4412126.

43. Kinkar L, Gasser RB, Webster BL, Rollinson D, Littlewood DTJ, Chang BCH, Stroehlein AJ, Korhonen PK, Young ND. Nanopore Sequencing Resolves Elusive Long Tandem-Repeat Regions in Mitochondrial Genomes. Int J Mol Sci. 2021 Feb 11;22(4):1811. doi: 10.3390/ijms22041811. PMID: 33670420; PMCID: PMC7918261.

44. Kluyver, T., Ragan-Kelley, B., Fernando Perez, Granger, B., Bussonnier, M., Frederic, J., Willing, C. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In F. Loizides & B. Schmidt (Eds.), *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90).

45. Kot W, Olsen NS, Nielsen TK, Hutinet G, de Crécy-Lagard V, Cui L, Dedon PC, Carstens AB, Moineau S, Swairjo MA, Hansen LH. Detection of preQ0 deazaguanine modifications in bacteriophage CAjan DNA using Nanopore sequencing reveals same hypermodification at two distinct DNA motifs. Nucleic Acids Res. 2020 Oct 9;48(18):10383-10396. doi: 10.1093/nar/gkaa735. PMID: 32941607; PMCID: PMC7544227.

46. Kumar S, Chinnusamy V, Mohapatra T. Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond. Front Genet. 2018 Dec 18;9:640. doi: 10.3389/fgene.2018.00640. PMID: 30619465; PMCID: PMC6305559.

47. Lee BK, Bhinge A, Iyer V, Wide-ranging functions of E2F4 in transcriptional activation and repression revealed by genome-wide analysis, Nucleic Acids Research, Volume 39, Issue 9, 1 May 2011, Pages 3558–3573, https://doi.org/10.1093/nar/gkq1313

48. Li JM, Hu PP, Shen X, Yu Y, Wang XF. E2F4-RB and E2F4-p107 complexes suppress gene expression by transforming growth factor beta through E2F binding sites. Proc Natl Acad Sci U S A. 1997 May 13;94(10):4948-53. doi: 10.1073/pnas.94.10.4948. PMID: 9144170; PMCID: PMC24611.

49. Li, L., Jin, M., Sun, C. et al. High Efficiency Hydrodynamic DNA Fragmentation in a Bubbling System. Sci Rep 7, 40745 (2017). https://doi.org/10.1038/srep40745

50. Lilue J, Doran AG, Fiddes IT. et. al. Sixteen diverse laboratory mouse reference genomes define strain-specific haplotypes and novel functional loci. Nat Genet. 2018 Nov;50(11):1574-1583. doi: 10.1038/s41588-018-0223-8. Epub 2018 Oct 1. PMID: 30275530; PMCID: PMC6205630.

51. Liu L, Li Y, Li T, Xie J, Chen C, Liu Q, Zhang S, Wu HC. Selective Detection of 8-Oxo-2'-deoxyguanosine in Single-Stranded DNA via Nanopore Sensing Approach. Anal Chem. 2016 Jan 19;88(2):1073-7. doi: 10.1021/acs.analchem.5b04102. Epub 2015 Dec 29. PMID: 26699617.

52. Liu, Y., Rosikiewicz, W., Pan, Z. et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. Genome Biol 22, 295 (2021). https://doi.org/10.1186/s13059-021-02510-z

53. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nat Rev Genet. 2020 Oct;21(10):597-614. doi: 10.1038/s41576-020-0236-x. Epub 2020 Jun 5. PMID: 32504078; PMCID: PMC7877196.

54. Luo GZ, Blanco MA, Greer EL, He C, Shi Y. DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes? Nat Rev Mol Cell Biol. 2015 Dec;16(12):705-10. doi: 10.1038/nrm4076. Epub 2015 Oct 28. PMID: 26507168; PMCID: PMC4763336.

55. Lüth, T., Wasner, L., Klein, C., Schaake, S., Tse, R., Pereira, S., Laß, J., Sinkkonen, L., Grünewald, A., Trinh, J. (2021). Nanopore Single-Molecule Sequencing for Mitochondrial DNA Methylation Analysis: Investigating Parkin-Associated Parkinsonism as a Proof of Concept. Frontiers in Aging Neuroscience. 13. 713084. 10.3389/fnagi.2021.713084.

56. Maejima Y, Sadoshima J. SUMOylation: a novel protein quality control modifier in the heart. Circ Res. 2014 Sep 26;115(8):686-9. doi: 10.1161/CIRCRESAHA.114.304989. PMID: 25258400; PMCID: PMC4181369.

57. Mastroeni D, Khdour OM, Delvaux E, Nolz J, Olsen G, Berchtold N, Cotman C, Hecht SM, Coleman PD. Nuclear but not mitochondrial-encoded oxidative phosphorylation genes are altered in aging, mild cognitive impairment, and Alzheimer's disease. Alzheimers Dement. 2017 May;13(5):510-519. doi: 10.1016/j.jalz.2016.09.003. Epub 2016 Oct 25. PMID: 27793643; PMCID: PMC5967608.

58. Menga A, Palmieri EM, Cianciulli A, Infantino V, Mazzone M, Scilimati A, Palmieri F, Castegna A, Iacobazzi V. SLC25A26 overexpression impairs cell function via mtDNA hypermethylation and rewiring of methyl metabolism. FEBS J. 2017 Mar;284(6):967-984. doi: 10.1111/febs.14028. Epub 2017 Feb 21. PMID: 28118529.

59. Milosevic, M., Arsic, A., Cvetkovic, Z., Vucic, V. (2021). Memorable Food: Fighting Age-Related Neurodegeneration by Precision Nutrition. Frontiers in Nutrition. 8. 10.3389/fnut.2021.688086.

60. Montoya J, Christianson T, Levens D, Rabinowitz M, Attardi G. Identification of initiation sites for heavy-strand and light-strand transcription in human mitochondrial DNA. Proc Natl Acad Sci U S A. 1982 Dec;79(23):7195-9. doi: 10.1073/pnas.79.23.7195. PMID: 6185947; PMCID: PMC347305.

61. Montoya J, Gaines GL, Attardi G. The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. Cell. 1983 Aug;34(1):151-9. doi: 10.1016/0092-8674(83)90145-9. PMID: 6883508.

62. Mookerjee SA, Gerencser AA, Nicholls DG, Brand MD. Quantifying intracellular rates of glycolytic and oxidative ATP production and consumption using extracellular flux measurements. J Biol Chem. 2017 Apr 28;292(17):7189-7207. doi: 10.1074/jbc.M116.774471. Epub 2017 Mar 7. Erratum in: J Biol Chem. 2018 Aug 10;293(32):12649-12652. PMID: 28270511; PMCID: PMC5409486.

63. Moore, L., Le, T. & Fan, G. DNA Methylation and Its Basic Function. Neuropsychopharmacol 38, 23–38 (2013). https://doi.org/10.1038/npp.2012.112

64. Moreno-Loshuertos R, Movilla N, Marco-Brualla J, Soler-Agesta R, Ferreira P, Enríquez JA, Fernández-Silva P. A Mutation in Mouse MT-ATP6 Gene Induces Respiration Defects and Opposed Effects on the Cell Tumorigenic Phenotype. Int J Mol Sci. 2023 Jan 9;24(2):1300. doi: 10.3390/ijms24021300. PMID: 36674816; PMCID: PMC9865613.

65. Morin AL, Win PW, Lin AZ, Castellani CA. Mitochondrial genomic integrity and the nuclear epigenome in health and disease. Front Endocrinol (Lausanne). 2022 Nov 7;13:1059085. doi: 10.3389/fendo.2022.1059085. PMID: 36419771; PMCID: PMC9678080.

66. Munchel S, Hoang Y, Zhao Y, Cottrell J, Klotzle B, Godwin AK, Koestler D, Beyerlein P, Fan JB, Bibikova M, Chien J. Targeted or whole genome sequencing of formalin fixed tissue samples: potential applications in cancer genomics. Oncotarget. 2015 Sep 22;6(28):25943-61. doi: 10.18632/oncotarget.4671. PMID: 26305677; PMCID: PMC4694877.

67. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2023 Jun 11]. Available from: https://www.ncbi.nlm.nih.gov/

68. Nicholls TJ, Minczuk M. In D-loop: 40 years of mitochondrial 7S DNA. Exp Gerontol. 2014 Aug;56:175-81. doi: 10.1016/j.exger.2014.03.027. Epub 2014 Apr 4. PMID: 24709344.

69. Ouyang Y, Wu Q, Li J, Sun S, Sun S. S-adenosylmethionine: A metabolite critical to the regulation of autophagy. Cell Prolif. 2020 Nov;53(11):e12891. doi: 10.1111/cpr.12891. Epub 2020 Oct 8. PMID: 33030764; PMCID: PMC7653241.

70. Oxford Nanopore (2008). *Epi2ME*

71. Oxford Nanopore (2023). *Epi2ME Labs*

72. Oxford Nanopore 2021, Nanopore protocol Rapid Sequencing (SQK-RAD004), accessed February 12th, 2023 https://sites.dundee.ac.uk/dundee-daffodil/wp-content/uploads/sites/190/2022/03/rapid-sequencing-sqk-rad004-RSE_9046_v1_revY_14Aug2019-minion.pdf

73. Oxford Nanopore 2023 DOCUMENTATION/SEQUENCE/SEQUENCING SOFTWARE/GUPPY PROTOCOL, Oxford Nanopore Technologies, accessed February 11th, 2024 https://community.nanoporetech.com/docs/prepare/library_prep_protocols/Guppy-protocol/v/gpb_2003_v1_revax_14dec2018/setting-up-a-run-configurations-and-parameters

74. Oxford Nanopore 2023 How Basecalling Works, Oxford Nanopore Technologies, accessed October 10th, 2023 https://nanoporetech.com/how-it-works/basecalling

75. Oxford Nanopore 2023 MinION Mk1C Brochure, Oxford Nanopore Technologies, accessed October 10th, 2023 https://nanoporetech.com/sites/default/files/s3/literature/MinION-Mk1C-brochure.pdf

76. Oxford Nanopore 2023 PromethION 2 Solo Brochure, Oxford Nanopore Technologies, accessed October 10th, 2023 https://nanoporetech.com/sites/default/files/s3/literature/J000949_PromethION%20P2%20Brochure_DIGITAL.pdf

77. Pakula A, Spinazzola JM, Gussoni E. Purification of Myogenic Progenitors from Human Muscle Using Fluorescence-Activated Cell Sorting (FACS). Methods Mol Biol. 2019;1889:1-15. doi: 10.1007/978-1-4939-8897-6_1. PMID: 30367405; PMCID: PMC6472699.

78. Park Y, Wu H (2016). "Differential methylation analysis for BS-seq data under general experimental design." Bioinformatics. doi:10.1093/bioinformatics/btw026.

79. Patil V, Cuenin C, Chung F, Aguilera JRR, Fernandez-Jimenez N, Romero-Garmendia I, Bilbao JR, Cahais V, Rothwell J, Herceg Z. Human mitochondrial DNA is extensively methylated in a non-CpG context. Nucleic Acids Res. 2019 Nov 4;47(19):10072-10085. doi: 10.1093/nar/gkz762. PMID: 31665742; PMCID: PMC6821263.

80. Pramanik, Parthasarathi & Sarma, G.A.S.K & Murty, Om. (2006). Effect of DNA sample collection , storage, transport on DNA fingerprinting. International Journal of Medical Toxicology and Legal Medicine. 8. 11-3.

81. Qiagen 2024, QIAcube Connect product page, Qiagen, accessed February 12th, 2024 https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/instruments-equipment/qiacube-connect

82. Quail MA, Swerdlow H, Turner DJ. Improved protocols for the illumina genome analyzer sequencing system. Curr Protoc Hum Genet. 2009 Jul;Chapter 18:Unit 18.2. doi: 10.1002/0471142905.hg1802s62. PMID: 19582764; PMCID: PMC3849550.

83. Quispe-Tintaya W, White RR, Popov VN, Vijg J, Maslov AY. Fast mitochondrial DNA isolation from mammalian cells for next-generation sequencing. Biotechniques. 2013 Sep;55(3):133-6. doi: 10.2144/000114077. PMID: 24003945; PMCID: PMC4353588.

84. Rand AC, Jain M, Eizenga JM, Musselman-Brown A, Olsen HE, Akeson M, Paten B. Mapping DNA methylation with high-throughput nanopore sequencing. Nat Methods. 2017 Apr;14(4):411-413. doi: 10.1038/nmeth.4189. Epub 2017 Feb 20. PMID: 28218897; PMCID: PMC5704956.

85. Razin A, Cedar H. DNA methylation and gene expression. Microbiol Rev. 1991 Sep;55(3):451-8. doi: 10.1128/mr.55.3.451-458.1991. PMID: 1943996; PMCID: PMC372829

86. Robanus-Maandag E, Dekker M, van der Valk M, Carrozza ML, Jeanny JC, Dannenberg JH, Berns A, te Riele H. p107 is a suppressor of retinoblastoma development in pRb-deficient mice. Genes Dev. 1998 Jun 1;12(11):1599-609. doi: 10.1101/gad.12.11.1599. PMID: 9620848; PMCID: PMC316874.

87. Robertson, K., Ait-Si-Ali, S., Yokochi, T. et al. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet 25, 338–342 (2000).

88. Roger AJ, Muñoz-Gómez SA, Kamikawa R. The Origin and Diversification of Mitochondria. Curr Biol. 2017 Nov 6;27(21):R1177-R1192. doi: 10.1016/j.cub.2017.09.015. PMID: 29112874.

89. Romero-Garcia S, Prado-Garcia H, Carlos-Reyes A. Role of DNA Methylation in the Resistance to Therapy in Solid Tumors. Front Oncol. 2020 Aug 7;10:1152. doi: 10.3389/fonc.2020.01152. PMID: 32850327; PMCID: PMC7426728.

90. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

91. Runtuwene LR, Tuda JSB, Mongan AE, Suzuki Y. On-Site MinION Sequencing. Adv Exp Med Biol. 2019;1129:143-150. doi: 10.1007/978-981-13-6037-4_10. PMID: 30968366.

92. Saini, S.K., Mangalhara, K.C., Prakasam, G. et al. DNA Methyltransferase1 (DNMT1) Isoform3 methylates mitochondrial genome and modulates its biology. Sci Rep 7, 1525 (2017).

93. Sanyal T, Bhattacharjee P, Bhattacharjee S, Bhattacharjee P. Hypomethylation of mitochondrial D-loop and ND6 with increased mitochondrial DNA copy number in the arsenic-exposed population. Toxicology. 2018 Sep 1;408:54-61.

94. Sharma H, Singh A, Sharma C, Jain SK, Singh N. Mutations in the mitochondrial DNA D-loop region are frequent in cervical cancer. Cancer Cell Int. 2005 Dec 16;5:34. doi: 10.1186/1475-2867-5-34. PMID: 16359547; PMCID: PMC1352382.

95. Sharma N, Pasala MS, Prakash A. Mitochondrial DNA: Epigenetics and environment. Environ Mol Mutagen. 2019 Oct;60(8):668-682. doi: 10.1002/em.22319. Epub 2019 Aug 6. PMID: 31335990; PMCID: PMC6941438.

96. Shi L, Tu BP. Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. Curr Opin Cell Biol. 2015 Apr;33:125-31. doi: 10.1016/j.ceb.2015.02.003. Epub 2015 Feb 20. PMID: 25703630; PMCID: PMC4380630.

97. Shumay E, Fowler JS. Identification and characterization of putative methylation targets in the MAOA locus using bioinformatic approaches. Epigenetics. 2010 May 16;5(4):325-42. doi: 10.4161/epi.5.4.11719. Epub 2010 May 5. PMID: 20421737; PMCID: PMC3169210.

98. Sirard, MA. Distribution and dynamics of mitochondrial DNA methylation in oocytes, embryos and granulosa cells. Sci Rep 9, 11937 (2019). https://doi.org/10.1038/s41598-019-48422-8

99. Smiraglia D, Kulawiec M, Bistulfi GL, Ghoshal A, and Singh KK (2008) A novel role for mitochondria in regulating epigenetic modifications in the nucleus, Cancer Biology & Therapy, 7:8, 1182-1190, DOI: 10.4161/cbt.7.8.6215

100. Smits P, Mattijssen S, Morava E, van den Brand M, van den Brandt F, Wijburg F, Pruijn G, Smeitink J, Nijtmans L, Rodenburg R, van den Heuvel L. Functional consequences of mitochondrial tRNA Trp and tRNA Arg mutations causing combined OXPHOS defects. Eur J Hum Genet. 2010 Mar;18(3):324-9. doi: 10.1038/ejhg.2009.169. Epub 2009 Oct 7. PMID: 19809478; PMCID: PMC2987211.

101. Sodiq O. Waheed, Shobhit S. Chaturvedi, Tatyana G. Karabencheva-Christova, and Christo Z. Christov ACS Catalysis 2021 11 (7), 3877-3890 DOI: 10.1021/acscatal.0c05034

102. Sood AJ, Viner C, Hoffman MM. DNAmod: the DNA modification database. J Cheminform. 2019 Apr 23;11(1):30. doi: 10.1186/s13321-019-0349-4. PMID: 31016417; PMCID: PMC6478773.

103. Stoccoro, A., & Coppedè, F. (2021). Mitochondrial DNA Methylation and Human Diseases. International journal of molecular sciences, 22(9), 4594. https://doi.org/10.3390/ijms22094594

104. Sun, X., Wang, Z., Cong, X., Lv, Y., Li, Z., Rong, L., Yang, T., Yu, D.”Mitochondrial gene COX2 methylation and downregulation is a biomarker of aging in heart mesenchymal stem cells”. International Journal of Molecular Medicine 47.1 (2021): 161-170.

105. Taanman JW. The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta. 1999 Feb 9;1410(2):103-23. doi: 10.1016/s0005-2728(98)00161-3. PMID: 10076021.

106. Tang JX, Thompson K, Taylor RW, Oláhová M. Mitochondrial OXPHOS Biogenesis: Co-Regulation of Protein Synthesis, Import, and Assembly Pathways. Int J Mol Sci. 2020 May 28;21(11):3820. doi: 10.3390/ijms21113820. PMID: 32481479; PMCID: PMC7312649.

107. Thermo Scientific 2023 Mitochondria Isolation Kit for Cultured Cells Product Page, ThermoFisher, accessed October 10th, 2023 https://www.thermofisher.com/order/catalog/product/89874?gclid=Cj0KCQjw7JOpBhCfARIsAL3bobc3q8Ri2LYDsr88memcOe5ToGX6YaSq-UoAwr22pUTqtufs_PnWI9gaAsNoEALw_wcB&ef_id=Cj0KCQjw7JOpBhCfARIsAL3bobc3q8Ri2LYDsr88memcOe5ToGX6YaSq-UoAwr22pUTqtufs_PnWI9gaAsNoEALw_wcB:G:s&s_kwcid=AL!3652!3!386244973888!!!g!!!6617619314!79549275115&cid=bid_pca_ppf_r01_co_cp1359_pjt0000_bid00000_0se_gaw_dy_pur_con

108. Tremolizzo L, Carboni G, Ruzicka WB, Mitchell CP, Sugaya I, Tueting P, Sharma R, Grayson DR, Costa E, Guidotti A. An epigenetic mouse model for molecular and behavioral neuropathologies related to schizophrenia vulnerability. Proc Natl Acad Sci U S A. 2002 Dec 24;99(26):17095-100. doi: 10.1073/pnas.262658999. Epub 2002 Dec 12. PMID: 12481028; PMCID: PMC139275.

109. Tubiana M. Tumor cell proliferation kinetics and tumor growth rate. Acta Oncol. 1989;28(1):113-21. Doi: 10.3109/02841868909111193. PMID: 2650719.

110. Ummarino S, Hausman C, Gaggi G, Rinaldi L, Bassal MA, Zhang Y, Seelam AJ, Kobayashi IS, Borchiellini M, Ebralidze AK, Ghinassi B, Trinh BQ, Kobayashi SS, Di Ruscio A. NAD Modulates DNA Methylation and Cell Differentiation. Cells. 2021 Nov 2;10(11):2986. doi: 10.3390/cells10112986. PMID: 34831209; PMCID: PMC8616462.

111. van der Wijst, M.G.P., van Tilburg, A.Y., Ruiters, M.H.J. et al. Experimental mitochondria-targeted DNA methylation, as potential regulator of mitochondrial gene expression. Sci Rep 7, 177 (2017).

112. Wanner, N., Larsen, P.A., McLain, A. *et al.* The mitochondrial genome and Epigenome of the Golden lion Tamarin from fecal DNA using Nanopore adaptive sequencing. *BMC Genomics* 22, 726 (2021). https://doi.org/10.1186/s12864-021-08046-7

113. Wilson DF. Oxidative phosphorylation: regulation and role in cellular and tissue metabolism. J Physiol. 2017 Dec 1;595(23):7023-7038. doi: 10.1113/JP273839. Epub 2017 Oct 29. PMID: 29023737; PMCID: PMC5709332.

114. Wirt SE, Sage J. p107 in the public eye: an Rb understudy and more. Cell Div. 2010 Apr 2;5:9. doi: 10.1186/1747-1028-5-9. PMID: 20359370; PMCID: PMC2861648.

115. Wong M, Gertz B, Chestnut BA, Martin LJ. Mitochondrial DNMT3A and DNA methylation in skeletal muscle and CNS of transgenic mouse models of ALS. Front Cell Neurosci. 2013 Dec 25;7:279. doi: 10.3389/fncel.2013.00279. PMID: 24399935; PMCID: PMC3872319.

116. Wright, D.J., Hall, N.A.L., Irish, N. et al. Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. BMC Genomics 23, 42 (2022). https://doi.org/10.1186/s12864-021-08261-2

117. Xu S, Zhang X, Liu C, Liu Q, Chai H, Luo Y, Li S. Role of Mitochondria in Neurodegenerative Diseases: From an Epigenetic Perspective. Front Cell Dev Biol. 2021 Aug 27;9:688789. doi: 10.3389/fcell.2021.688789. PMID: 34513831; PMCID: PMC8429841.

118. Xu, B., Li, X., Du, M. et al. Novel mutation of ND4 gene identified by targeted next-generation sequencing in patient with Leigh syndrome. J Hum Genet 62, 291–297 (2017). https://doi.org/10.1038/jhg.2016.127

119. Yaffe D, Saxel O. Serial passaging and differentiation of myogenic cells isolated from dystrophic mouse muscle. Nature. 1977 Dec 22-29;270(5639):725-7. doi: 10.1038/270725a0. PMID: 563524.

120. Zhao, L.N., Björklund, M., Caldez, M.J. et al. Therapeutic targeting of the mitochondrial one-carbon pathway: perspectives, pitfalls, and potential. Oncogene 40, 2339–2354 (2021). https://doi.org/10.1038/s41388-021-01695-8

121. Zuo T, Tycko B, Liu TM, Lin JJ, Huang TH. Methods in DNA methylation profiling. Epigenomics. 2009 Dec;1(2):331-45. doi: 10.2217/epi.09.31. PMID: 20526417; PMCID: PMC2880494

Appendix:

Appendix 1. differentially methylated sites in *MA* and *MD*. This table displays the genes that exhibited differential methylation, the positions along the mitochondrial genome where that differential methylation was found, the mean methylation frequency for each differentially methylated site in the p107KO *MA*, the mean methylation frequency for each differentially methylated site in the p107WT *MD*, the difference in methylation levels for each site between the samples, and the p-value for differential methylation at each site.

Gene	Position	p107KO Methylation	p107WT Methylation	Methylation Difference	p-value
ND2	4920	0.007496246	0.1458576	-0.1383613	4.307112e-11
ND2	4921	0.007496246	0.1458576	-0.1383613	4.307112e-11
ND2	4924	0.007496246	0.1458576	-0.1383613	4.307112e-11
ND2	4925	0.007496246	0.1458576	-0.1383613	4.307112e-11
ND2	4926	0.007496246	0.1458576	-0.1383613	4.307112e-11
ND2	4927	0.007496246	0.1458576	-0.1383613	4.307112e-11
ND2	4930	0.007496246	0.1458576	-0.1383613	4.278481e-11
ND2	4931	0.007496246	0.1458576	-0.1383613	4.307112e-11
ND2	4932	0.007496246	0.1458576	-0.1383613	4.307112e-11
ND2	4933	0.007456076	0.1457126	-0.1382565	4.240622e-11
ND2	4934	0.007416373	0.1455686	-0.1381522	4.038714e-11

ND2	4935	0.008135183	0.1472165	-0.1390814	1.501640e-11
ND2	4936	0.008337251	0.1419359	-0.1335986	1.919739e-11
tRNA-trp	4939	0.008293697	0.1418142	-0.1335205	1.908292e-11
tRNA-trp	4946	0.008166098	0.1414533	-0.1332872	1.652347e-11
tRNA-trp	4947	0.008216112	0.1415729	-0.1333568	1.746227e-11
tRNA-trp	4950	0.008224219	0.1415729	-0.1333487	1.744996e-11
tRNA-trp	4953	0.008182085	0.1414533	-0.1332712	1.662781e-11
tRNA-trp	4958	0.008188820	0.1414533	-0.1332645	1.673528e-11
tRNA-trp	4959	0.008146575	0.1413345	-0.1331879	1.594296e-11
tRNA-trp	4965	0.008161524	0.1413345	-0.1331729	1.629507e-11
tRNA-trp	4968	0.008175252	0.1413345	-0.1331592	1.636417e-11
ND4	11274	0.010306816	0.1432261	-0.1329193	6.214930e-13

ND4	11373	0.009538956	0.1435635	-0.1340246	3.424115e-11
ND4	11375	0.009650379	0.1438204	-0.1341701	3.332168e-11
ND4	11376	0.009707218	0.1439501	-0.1342428	3.499951e-11
ND4	11383	0.009802216	0.1443438	-0.1345416	3.964636e-11
ND4	11385	0.009861642	0.1444767	-0.1346150	4.165278e-11
ND4	11392	0.010108026	0.1450166	-0.1349086	5.063422e-11
ND4	11403	0.010285935	0.1455703	-0.1352843	6.000305e-11
ND4	11404	0.010353278	0.1457109	-0.1353576	7.215745e-11
ND4	11406	0.010421696	0.1458525	-0.1354308	6.615304e-11
ND4	11407	0.010421696	0.1458525	-0.1354308	7.583914e-11
ND4	11408	0.010421696	0.1458525	-0.1354308	6.615304e-11
ND4	11415	0.010702770	0.1464278	-0.1357251	7.998353e-11
ND4	11416	0.010702770	0.1464278	-0.1357251	8.024981e-11
ND4	11420	0.009972672	0.1536263	-0.1436537	1.444549e-10
ND4	11440	0.006847082	0.1498393	-0.1429922	1.517069e-08

ND4	11441	0.006847082	0.1498393	-0.1429922	1.517069e-08
ND4	11443	0.006847082	0.1498393	-0.1429922	1.517069e-08
ND4	11444	0.006847082	0.1498393	-0.1429922	1.517069e-08
ND4	11447	0.006897255	0.1500634	-0.1431661	1.596911e-08
ND4	11448	0.006948044	0.1502896	-0.1433416	1.680824e-08
ND4	11449	0.006999651	0.1505179	-0.1435183	1.769149e-08
ND4	11451	0.007105389	0.1509811	-0.1438757	1.959892e-08
ND5	12778	0.014071638	0.1405928	-0.1265212	8.093384e-16
ND5	12782	0.014141811	0.1406670	-0.1265252	9.265529e-16
ND5	12786	0.014420942	0.1409669	-0.1265459	1.113356e-15
ND5	12787	0.014492520	0.1410425	-0.1265500	1.207009e-15
ND5	12788	0.014426220	0.1409669	-0.1265406	1.150319e-15
ND5	12790	0.014337732	0.1410425	-0.1267048	1.085412e-15
ND5	12857	0.015399096	0.1402585	-0.1248594	4.101372e-17
ND5	12860	0.015259194	0.1401263	-0.1248671	5.937120e-17

Appendix 2. differentially methylated sites in *MC* and *MD*. This table displays the genes that exhibited differential methylation, the positions along the mitochondrial genome where that differential methylation was found, the mean methylation frequency for each differentially methylated site in the p107KO *MC*, the mean methylation frequency for each differentially methylated site in the p107WT *MD*, the difference in methylation levels for each site between the samples, and the p-value for differential methylation at each site.

Gene	Position	p107KO Methylation	p107WT Methylation	Methylation Difference	p-value
tRNA-Gln	3820	0.005798762	0.1333377	-0.1275390	4.468921e-16
tRNA-Gln	3821	0.005844863	0.1341565	-0.1283116	7.337090e-16
tRNA-Gln	3824	0.005868052	0.1342235	-0.1283555	7.582039e-16
tRNA-Gln	3825	0.005821489	0.1341565	-0.1283350	7.175426e-16
tRNA-Gln	3830	0.005767379	0.1340232	-0.1282559	6.644621e-16
tRNA-Gln	3832	0.005774447	0.1340232	-0.1282488	6.683320e-16
tRNA-Gln	3833	0.005797048	0.1340897	-0.1282927	6.912954e-16
tRNA-Trp	4920	0.006330863	0.1458576	-0.1395267	3.595070e-11
tRNA-Trp	4921	0.006330863	0.1458576	-0.1395267	3.595070e-11

tRNA-Trp	4924	0.006330863	0.1458576	-0.1395267	3.595070e-11
tRNA-Trp	4925	0.006330863	0.1458576	-0.1395267	3.595070e-11
tRNA-Trp	4926	0.006330863	0.1458576	-0.1395267	3.595070e-11
tRNA-Trp	4927	0.006330863	0.1458576	-0.1395267	3.595070e-11
tRNA-Trp	4930	0.006330863	0.1458576	-0.1395267	4.398429e-11
tRNA-Trp	4931	0.006330863	0.1458576	-0.1395267	3.595070e-11
tRNA-Trp	4932	0.006330863	0.1458576	-0.1395267	3.595070e-11
tRNA-Trp	4933	0.006289065	0.1457126	-0.1394235	3.525173e-11
tRNA-Trp	4934	0.006247861	0.1455686	-0.1393207	3.344016e-11
tRNA-Trp	4935	0.008316195	0.1472165	-0.1389004	2.156452e-11
tRNA-Trp	4936	0.008261304	0.1419359	-0.1336746	2.553670e-11

tRNA-Trp	4939	0.008207178	0.1418142	-0.1336070	2.517220e-11
tRNA-Trp	4946	0.008056001	0.1414533	-0.1333973	2.141060e-11
tRNA-Trp	4947	0.008093698	0.1415729	-0.1334792	2.254446e-11
tRNA-Trp	4950	0.008079427	0.1415729	-0.1334934	2.958331e-11
tRNA-Trp	4953	0.008027784	0.1414533	-0.1334255	2.793614e-11
tRNA-Trp	4958	0.008034088	0.1414533	-0.1334192	2.813870e-11
tRNA-Trp	4959	0.007989724	0.1413345	-0.1333447	2.668016e-11
tRNA-Trp	4965	0.007975246	0.1413345	-0.1333592	2.032033e-11
tRNA-Trp	4968	0.007980883	0.1413345	-0.1333536	2.689977e-11
COX1	6482	0.007150591	0.1360740	-0.1289234	8.081386e-16
COX1	6484	0.007146378	0.1360740	-0.1289276	8.066096e-16
COX1	6488	0.007182181	0.1361435	-0.1289613	8.462047e-16

COX1	6489	0.007218369	0.1362133	-0.1289950	8.877210e-16
COX1	6493	0.007218369	0.1362133	-0.1289950	8.877210e-16
COX1	6497	0.007218369	0.1362133	-0.1289950	8.892321e-16
COX1	6498	0.007182181	0.1361435	-0.1289613	8.483425e-16
COX1	6500	0.007177931	0.1361435	-0.1289656	8.461802e-16
COX1	6501	0.007181853	0.1361435	-0.1289617	8.493636e-16
COX1	6506	0.007149939	0.1360740	-0.1289241	8.134784e-16
COX1	6509	0.007118325	0.1360047	-0.1288864	7.786202e-16
COX1	6511	0.007118325	0.1360047	-0.1288864	7.786202e-16
COX1	6513	0.007153829	0.1358819	-0.1287281	2.466741e-15
mtATP6	8379	0.005653940	0.1400831	-0.1344291	6.923565e-11
mtATP6	8382	0.005629985	0.1399597	-0.1343298	6.569866e-11
mtATP6	8383	0.005653942	0.1400831	-0.1344291	6.911570e-11
mtATP6	8387	0.005638970	0.1400831	-0.1344441	7.428657e-11
mtATP6	8388	0.005638970	0.1400831	-0.1344441	7.428657e-11

mtATP6	8389	0.005638970	0.1400831	-0.1344441	6.859400e-11
mtATP6	8394	0.005620848	0.1399597	-0.1343389	7.066411e-11
mtATP6	8395	0.005620848	0.1399597	-0.1343389	6.524415e-11
mtATP6	8397	0.005608093	0.1399597	-0.1343517	7.015396e-11
mtATP6	8399	0.005611050	0.1399597	-0.1343487	6.474815e-11
mtATP6	8402	0.005574851	0.1398372	-0.1342624	6.615320e-11
mtATP6	8403	0.005574851	0.1398372	-0.1342624	6.113897e-11
mtATP6	8407	0.005554501	0.1397155	-0.1341610	5.806018e-11
mtATP6	8409	0.005554501	0.1397155	-0.1341610	6.282056e-11
mtATP6	8410	0.005554501	0.1397155	-0.1341610	5.806018e-11
mtATP6	8414	0.005562549	0.1397155	-0.1341529	5.818341e-11
mtATP6	8419	0.004932973	0.1399597	-0.1350268	5.016824e-11
mtATP6	8421	0.004956280	0.1400831	-0.1351268	5.668724e-11
mtATP6	8422	0.004956280	0.1400831	-0.1351268	5.278194e-11
mtATP6	8424	0.004956280	0.1400831	-0.1351268	5.668724e-11

mtATP6	8425	0.004956280	0.1400831	-0.1351268	5.668724e-11
mtATP6	8427	0.004979830	0.1402072	-0.1352274	5.552695e-11
mtATP6	8429	0.005027675	0.1404578	-0.1354302	6.144471e-11
mtATP6	8439	0.005060102	0.1404578	-0.1353977	6.179268e-11
mtATP6	8441	0.005065827	0.1404578	-0.1353920	6.180030e-11
mtATP6	8442	0.005090521	0.1405844	-0.1354939	6.499151e-11
mtATP6	8449	0.005026103	0.1403321	-0.1353060	6.217931e-11
mtATP6	8451	0.005026103	0.1403321	-0.1353060	6.217931e-11
mtATP6	8452	0.005026103	0.1403321	-0.1353060	6.217931e-11
mtATP6	8454	0.005050372	0.1404578	-0.1354075	6.535410e-11
mtATP6	8455	0.005050372	0.1404578	-0.1354075	6.074286e-11
mtATP6	8458	0.005067773	0.1405844	-0.1355166	6.338879e-11
mtATP6	8460	0.005092489	0.1407118	-0.1356193	6.660124e-11
mtATP6	8463	0.005079252	0.1405844	-0.1355051	6.806395e-11
mtATP6	8479	0.005146862	0.1408400	-0.1356931	7.489813e-11

mtATP6	8480	0.005172417	0.1409691	-0.1357966	7.297401e-11
mtATP6	8481	0.005172417	0.1409691	-0.1357966	7.297401e-11
mtATP6	8482	0.005172417	0.1409691	-0.1357966	7.297401e-11
mtATP6	8484	0.005203409	0.1410990	-0.1358956	7.679581e-11
mtATP6	8485	0.005203409	0.1410990	-0.1358956	7.679581e-11
mtATP6	8488	0.005144232	0.1407118	-0.1355675	6.595002e-11
mtATP6	8490	0.005200693	0.1409691	-0.1357684	7.866417e-11
mtATP6	8491	0.005193210	0.1409691	-0.1357758	7.228323e-11
mtATP6	8494	0.005121166	0.1405844	-0.1354632	6.170016e-11
mtATP6	8495	0.005146431	0.1407118	-0.1355653	6.478509e-11
mtATP6	8500	0.005158276	0.1407118	-0.1355535	6.452766e-11
ND4	11420	0.010935655	0.1536263	-0.1426907	2.911941e-10
ND5	12778	0.008734766	0.1405928	-0.1318580	6.298434e-17
ND5	12782	0.008758536	0.1406670	-0.1319085	7.082660e-17
ND5	12786	0.008904174	0.1409669	-0.1320627	8.332824e-17

ND5	12787	0.008941389	0.1410425	-0.1321011	8.676882e-17
ND5	12788	0.008891854	0.1409669	-0.1320750	8.257566e-17
ND5	12790	0.008928963	0.1410425	-0.1321136	8.596752e-17
ND5	12857	0.011937868	0.1402585	-0.1283206	1.156172e-17
ND5	12860	0.011833290	0.1401263	-0.1282930	1.427328e-17
ND5	12867	0.011786964	0.1374022	-0.1256153	1.614752e-17
ND5	12880	0.011647302	0.1367767	-0.1251294	3.921386e-17
ND5	12882	0.011641931	0.1367767	-0.1251347	3.911783e-17
ND5	12884	0.011514884	0.1366567	-0.1251418	3.469131e-17
ND5	12885	0.011559166	0.1367166	-0.1251574	3.627198e-17
ND5	12890	0.012071573	0.1366567	-0.1245852	5.106800e-17
ND5	12891	0.012020075	0.1365971	-0.1245770	4.865336e-17
ND5	12895	0.012014633	0.1365971	-0.1245824	4.849366e-17
ND5	12896	0.011963635	0.1365376	-0.1245740	4.620136e-17
ND5	12900	0.012101420	0.1367166	-0.1246152	5.279130e-17

ND5	12902	0.012101420	0.1367166	-0.1246152	5.279130e-17
ND5	12903	0.012148086	0.1367767	-0.1246286	5.518495e-17
ND5	12904	0.012148086	0.1367767	-0.1246286	5.518495e-17
ND5	12907	0.012142529	0.1367767	-0.1246341	5.495405e-17
ND5	12910	0.012084893	0.1367166	-0.1246317	5.206146e-17
ND5	12912	0.012079400	0.1367166	-0.1246372	5.179713e-17
ND5	12917	0.012048918	0.1367166	-0.1246677	5.034022e-17
ND5	12919	0.012043458	0.1367166	-0.1246731	5.007387e-17
ND5	12920	0.012043458	0.1367166	-0.1246731	5.007387e-17
ND5	12929	0.012018565	0.1367166	-0.1246980	4.880947e-17
ND5	12932	0.011916927	0.1365971	-0.1246801	4.425009e-17
ND5	12933	0.011916927	0.1365971	-0.1246801	4.425009e-17